Notas em Matemática Aplicada e-ISSN 2236-5915

Volume 42, 2012

Editores

Cassio Machiaveli Oishi
Universidade Estadual Paulista - UNESP
Presidente Prudente, SP, Brasil

Fernando Rodrigo Rafaeli
Universidade Estadual Paulista - UNESP
São José do Rio Preto, SP, Brasil

Rosana Sueli da Motta Jafelice (Editor Chefe)
Universidade Federal de Uberlândia - UFU
Uberlândia, MG, Brasil

Rubens de Figueiredo Camargo
Universidade Estadual Paulista - UNESP
Bauru, SP, Brasil

Sezimária de Fátima P. Saramago
Universidade Federal de Uberlândia - UFU
Uberlândia, MG, Brasil

Vanessa Avansini Botta Pirani (Editor Adjunto)
Universidade Estadual Paulista - UNESP
Presidente Prudente, SP, Brasil

SBMAC **Sociedade Brasileira de Matemática Aplicada e Computacional**

2012

A Sociedade Brasileira de Matemática Aplicada e Computacional - SBMAC publica, desde as primeiras edições do evento, monografias dos cursos que são ministrados nos CNMAC.

Para a comemoração dos 25 anos da SBMAC, que ocorreu durante o XXVI CNMAC em 2003, foi criada a série **Notas em Matemática Aplicada** para publicar as monografias dos minicursos ministrados nos CNMAC, o que permaneceu até o XXXIII CNMAC em 2010.

A partir de 2011, a série passa a publicar, também, livros nas áreas de interesse da SBMAC. Os autores que submeterem textos à série Notas em Matemática Aplicada devem estar cientes de que poderão ser convidados a ministrarem minicursos nos eventos patrocinados pela SBMAC, em especial nos CNMAC, sobre assunto a que se refere o texto.

O livro deve ser preparado em **Latex (compatível com o Miktex versão 2.7), as figuras em eps** e deve ter entre **80 e 150 páginas**. O texto deve ser redigido de forma clara, acompanhado de uma excelente revisão bibliográfica e de **exercícios de verificação de aprendizagem** ao final de cada capítulo.

Veja todos os títulos publicados nesta série na página
http://www.sbmac.org.br/notas.php

SBMAC **Sociedade Brasileira de Matemática Aplicada e Computacional**

2012

AVANÇOS EM MÉTODOS DE KRYLOV PARA SOLUÇÃO DE SISTEMAS LINEARES DE GRANDE PORTE

2ª edição

Luiz Mariano Carvalho
luizmc@ime.uerj.br

Departamento de Matemática Aplicada
Instituto de Matemática e Estatística
Universidade do Estado do Rio de Janeiro
Rio de Janeiro, Brasil

Serge Gratton
serge.gratton@enseeiht.fr

Institut National Polytechnique de Toulouse
European Centre for Research and Advanced Training in Scientific Computation - CERFACS
Toulouse, França

Sociedade Brasileira de Matemática Aplicada e Computacional

São Carlos - SP, Brasil
2012

Coordenação Editorial: Elbert Einstein Nehrer Macau

Coordenação Editorial da Série: Rosana Sueli da Motta Jafelice

Editora: SBMAC

Capa: Matheus Botossi Trindade

Patrocínio: SBMAC

Catalogação elaborada pela Biblioteca do IBILCE/UNESP
Bibliotecária: Maria Luiza Fernandes Jardim Froner

Carvalho, Luiz M.
Avanços em Métodos de Krylov para Solução de Sistemas Lineares de Grande Porte - São Carlos, SP : SBMAC, 2012, 159 p., 20.5 cm - (Notas em Matemática Aplicada; v. 42) - 2ª edição

e-ISBN 978-85-86883-68-2

1. Subespaços de Krylov 2. Precondicionadores
3. Métodos com Recomeço 4. Deflação
5. Precondicionadores Flexíveis 6. GMRES
I. Carvalho, Luiz M. II. Gratton, Serge. III. Título. IV. Série

CDD - 51

Conteúdo

Many of the scientific treatises of today are formulated in a half-mystical language, as though to impress the reader with the uncomfortable feeling that he is in the permanent presence of a superman. The present book is conceived in a humble spirit and is written for humble people. The author knows from past experience that one outstanding weakness of our present system of college education is the custom of classing certain fundamental and apparently simple concepts as "elementary," and of relegating them to an age-level at which the student's mind is not mature enough to grasp their true meaning. The fruits of this error can be observed daily. (Cornelius Lanczos, página ix no Prefácio de [75], de 1948).

Prefácio

O principal objetivo desse trabalho é apresentar alguns dos avanços recentes nos métodos de cálculo de soluções aproximadas de sistemas lineares de grande porte através de projeções em **subespaços de Krylov**. Os últimos 60 anos têm testemunhado um desenvolvimento constante desses métodos e, não por acaso, também dos computadores [103].

Não pode restar a menor dúvida de que, atualmente, o **melhor e mais utilizado método** para resolver sistemas lineares **é a eliminação gaussiana com pivoteamento parcial**. Mas a ordem das matrizes a serem resolvidas em problemas atuais alcançam cifras enormes. No caso de matrizes esparsas, sem uma estrutura conhecida, o procedimento padrão de eliminação gaussiana não é indicado, pois rapidamente chega-se à exaustão de memória e, em muitos casos, o tempo necessário para a solução é inviável. Nesse momento, os métodos iterativos são chamados à cena, e dentre eles, os mais utilizados, em aplicações acadêmicas e industriais, são os métodos de Krylov, por suas boas propriedades numéricas e computacionais.

No capítulo 1, apresentamos as definições e vários resultados essenciais para a os desenvolvimentos ulteriores. Sempre podemos interpretar os métodos de Krylov como **de projeção**, discutidos na seção 1.3. Entre as várias classificações existentes para os métodos de Krylov, uma delas os divide em duas grandes classes: a dos baseados no procedimento de **Arnoldi** e a dos métodos originários no procedimento de

Lanczos não-simétrico [101]. De acordo com ela, os métodos analisados nesse livro devem ser considerados como de Arnoldi. No capítulo 2, discutimos o procedimento de Arnoldi e dois representantes dessa classe: o **FOM** e o **GMRES**. Para permitir uma melhor compreensão desses métodos e da solução que eles fornecem, no capítulo 3, analisamos a qualidade dos **erros** da solução de sistemas lineares, **precondicionadores** e **critérios de parada**. A seguir, no capítulo 4, sistematizamos conceitos úteis à compreensão dos métodos de Krylov: os **pares harmônicos de Ritz**, os **métodos de ortogonalização** mais utilizados. Essas ferramentas teóricas ajudam a entender os métodos descritos no capítulo 5. Finalmente, no capítulo 6, fazemos o estudo de um novo método de Krylov que é baseado em algumas das ideias apresentadas nos capítulos anteriores.

Adotaremos em todo esse trabalho as convenções do programa Matlab® para fazer referência a vetores e matrizes, ou às suas partes. Além disso, usaremos letras maiúsculas em itálico para matrizes, A, letras maiúsculas cursivas para espaços vetoriais, $\mathcal{V}$, e letras minúsculas em itálico para vetores, x. Sendo assim $A(:, j)$ refere-se a $j-$ésima coluna da matriz A, $A(i, :)$ refere-se a sua $i-$ésima linha e $A(i, j)$ ao elemento da $i-$ésima linha e $j-$ésima coluna, podemos usar a_{ij} para representar a mesma informação. A não ser em casos especiais, nos referiremos sem distinção a uma transformação linear ou à matriz que a representa na base subjacente.

Os 95 exercícios são destinados ao aprofundamento e fixação das ideias, podendo ser demonstrações, desenvolvimento de programas ou ponteiros com referências para novos estudos.

Realizamos algumas modificações nessa segunda edição. Introduzimos uma revisão sistemática dos principais conceitos de álgebra linear necessários para a compreensão dos métodos; eles serão tratados no Apêndice A. Uma outra alteração é a seção 4.2, ela é dedicada ao estudo dos principais métodos para ortogonalização de bases de espaços vetoriais. Por fim, retiramos o estudo de métodos de Krylov em bloco por avaliarmos que o tratamento desse tópico precisa de mais espaço e de um maior detalhamento do que o disponível nessa obra. Um outro enfoque dos métodos de Krylov, a partir de conceitos geométricos,

pode ser visto em [31].

Por fim, ambos os autores gostariam de agradecer à **SBMAC** pela oportunidade de viabilizar a produção desse trabalho e, também, ao professor **Nelson Maculan** (UFRJ) pelo apoio decisivo a essa realização. O primeiro autor gostaria de agradecer à fraterna acolhida que recebeu no CERFACS por parte de **Xavier Vasseur** e de **Iain S. Duff** do **Parallel Algorithms Project do CERFACS**, Toulouse, França, aonde uma parte desse trabalho foi desenvolvida.

Rio de Janeiro, 27 de novembro de 2011
Luiz Mariano Carvalho

Toulouse, 27 de novembro de 2011
Serge Gratton

Capítulo 1

Espaços de Krylov e Métodos de Projeção

Apresentaremos e discutiremos propriedades dos espaços de Krylov[1], nas seções 1.1 e 1.2. Esses espaços apareceram originalmente em uma técnica proposta por A.N. Krylov para construção de polinômios característicos de matrizes. Sem entrar em detalhes (que podem ser vistos em [81, seção 7.11]), o resultado do método é a construção de matrizes K, **regular**[2], e H, Hessenberg, tal que o produto $K^{-1}AK = H$ seja válido. Trata-se de uma relação de similaridade e, portanto, A e H têm o mesmo polinômio característico. As colunas da matriz K, em uma primeira fase do método, são construídas pela multiplicação de um vetor b por A, a saber, a $j-$ésima coluna de $K(:, j)$ é dada por $A^{j-1}b$, ou seja

$$K = \begin{pmatrix} b & Ab & A^2b & \dots & A^{k-1}b. \end{pmatrix}$$

[1] Aleksei Nikolaevich Krylov (1863-1945) mostrou em 1931 [74] como usar sequências da forma $\{b, Ab, A^2b, \dots\}$ para construir o polinômio característico de uma matriz. Krylov foi um matemático aplicado russo (engenheiro marítimo de formação [nota dos tradutores]) cujos interesses científicos ultrapassavam em muito as áreas de seu treinamento inicial em ciência naval, que envolviam flutuação, estabilidade, etc. Krylov foi diretor do Instituto de Física-Matemática da Academia de Ciências da União Soviética de 1927 a 1932, e em 1943 ganhou um "prêmio estatal" por suas teorias sobre bússolas. Foi condecorado como "herói do trabalho socialista" e é um dos poucos matemáticos que tem um acidente geográfico lunar associado a seu nome, trata-se da cratera Krylov. (traduzido pelos autores de [81])

[2] Matriz é regular quando for não singular, ou seja, inversível.

A técnica desenvolvida nesse método está na gênese de todos os, assim chamados, métodos de Krylov.

Na seção 1.3, trataremos das propriedades de métodos de projeção genéricos. E por fim, sintetizaremos, com a combinação de métodos de projeção em subespaços de Krylov.

1.1 Espaços de Krylov

Chamaremos de **espaço de Krylov** o conjunto formado por todas combinações lineares dos vetores $\mathcal{K}(A, b) := \langle b, Ab, A^2b, \ldots\rangle$[3], chamaremos de **subespaço de Krylov** o conjunto formado pelas combinações lineares dos k vetores

$$\mathcal{K}(A, b, k) = \mathcal{K}_k(A, b) := \langle b, Ab, A^2b, \ldots, A^{k-1}b\rangle.$$

Uma **matriz de Krylov** é a matriz cujas colunas são os vetores que originam um subespaço de Krylov, e será denotada

$$K_k := \begin{pmatrix} b & Ab & A^2b & \ldots & A^{k-1}b. \end{pmatrix}$$

Uma **sequência de Krylov** é uma sequência de vetores $\begin{pmatrix} x_k \end{pmatrix} \in \mathbb{C}^m$ tal que $x_k = A^{k-1}b$. Uma sequência de Krylov pode, ou não, ser formada por vetores linearmente independentes, para tratar dessa propriedade, definimos um divisor particular do polinômio mínimo da matriz A.

Sejam $A \in \mathbb{C}^{m\times m}$ e $b \in \mathbb{C}^m$. O **polinômio mínimo de um vetor** b em relação à matriz A [68, pág. 18, seção 1.5] é o polinômio mônico de grau mínimo tal que

$$p(A)b = 0.$$

Se A^kb é o primeiro vetor que se torna uma combinação linear dos vetores anteriores de uma sequência de Krylov, ou seja,

$$A^kb = \sum_{i=0}^{k-1} \alpha_i A^i b,$$

[3]Usamos a notação $\langle u_1, u_2, \ldots, u_k\rangle$ para representar o **subespaço gerado** por todas das combinações lineares dos vetores u_i, $i = 1 : k$. E a notação := para dizer que a parte à direita desse símbolo é a definição do que se encontra à esquerda dele.

então, $p(x) = x^k - \sum_{i=0}^{k-1} \alpha_i x^i$ (ou $p(x) = 1$, quando $b = 0$) é o polinômio mínimo de b em relação a A.

Observação 1.1. *O polinômio mínimo de uma matriz A é o polinômio de menor grau que é o polinômio mínimo para todos os vetores do espaço vetorial considerado em relação à matriz A [68, pág. 18, seção 1.5].*

Tornando mais precisa a observação anterior, o resultado a seguir compara o polinômio mínimo de um vetor relativo a uma matriz, com o polinômio mínimo dessa matriz.

Teorema 1.1 ([81, pág. 647, seção 7.11])**.** *Sejam $A \in \mathbb{C}^{m\times m}$ e $V = (v_1, v_2, \ldots, v_m)$ uma base ordenada para $\mathbb{C}^m$. Se $p_i(t)$ é o polinômio mínimo de v_i em relação a A, então o polinômio mínimo de A, $q_A(t)$, é divisível por cada $p_i(t)$ e, dado um outro polinômio $p(t)$, caso cada $p_i(t)$ divida $p(t)$ então $q_A(t)$ também divide $p(t)$, ou seja, o polinômio mínimo de A é o mínimo múltiplo comum de todos os polinômios mínimos dos vetores de $\mathbb{C}^m$ em relação a A.*

Demonstração: Exercício 1.

Uma das principais motivações para os métodos iterativos de projeção em subespaços de Krylov é o seguinte teorema.

Teorema 1.2 ([70, Teorema 1])**.** *Sejam $A \in \mathbb{C}^{m\times m}$, matriz regular, e $b \in \mathbb{C}^m$. Seja x^* a solução exata do sistema linear $Ax = b$. Seja x_0 um valor inicial para x^* e $r_0 = b - Ax_0$, o resíduo inicial. Se o polinômio mínimo do vetor r_0 relativo à A tem grau $k-1$, então $x^* - x_0$ pertence ao espaço de Krylov $\mathcal{K}_{(k-1)}(A, r_0)$.*

Demonstração: Seja $p_{r_0}(t)$ o polinômio mínimo de r_0 em relação a A, logo

$$p_{r_0}(A)r_0 = \alpha_0 r_0 + \alpha_1 A r_0 + \alpha_2 A^2 r_0 + \ldots + \alpha_{k-1} A^{k-1} r_0 = 0.$$

Como A é regular então existe A^{-1} e temos

$$\alpha_0 A^{-1} r_0 + \alpha_1 r_0 + \alpha_2 A r_0 + \ldots + \alpha_{k-1} A^{k-2} r_0 = 0.$$

Pelo teorema 1.1, $p_{r_0}(t)$ divide o polinômio mínimo de A; sendo A regular, não possui o autovalor 0 e já que α_0 é o produto dos autovalores de A então, $\alpha_0 \neq 0$. Podemos escrever:

$$A^{-1}(b - Ax_0) = x^* - x_0 = -\frac{\alpha_1}{\alpha_0}r_0 - \frac{\alpha_2}{\alpha_0}Ar_0 - \ldots - \frac{\alpha_{k-1}}{\alpha_0}A^{k-2}r_0.$$

Logo $(x^* - x_0) \in \mathcal{K}_{(k-1)}(A, r_0)$. ■

Observação 1.2. *Se estamos em busca de uma solução para $Ax = b$, o espaço natural de busca é o subespaço de Krylov gerado por A e r_0. Segundo observam Ilpsen & Meyer em [70], o polinômio mínimo de um vetor em relação a uma matriz pode ter um grau bem menor do que o polinômio mínimo da mesma matriz. E por isso, dependendo de r_0, que por sua vez depende de x_0 e do lado direito b, um método de Krylov pode convergir em um número de passos notadamente inferior ao grau do polinômio mínimo da matriz em questão.*

Observação 1.3. *A versão do teorema 1.2 aqui apresentada é um pouco diferente em [70], pois estamos usando o polinômio mínimo de r_0 em relação a A e um valor inicial x_0, mas em essência é o mesmo resultado.*

Veremos, a seguir, algumas propriedades dos subespaços de Krylov.

Teorema 1.3 (Propriedades dos Subespaços de Krylov [120, pág. 267])**.** *Sejam $A \in \mathbb{C}^{m\times m}$ e $b \in \mathbb{C}^m$. Então*

1. *Uma sequência de subespaços de Krylov satisfaz*

$$\mathcal{K}_k(A, b) \subset \mathcal{K}_{k+1}(A, b)$$

e

$$A\mathcal{K}_k(A, b) \subset \mathcal{K}_{k+1}(A, b).$$

2. *Se $\alpha_i \in \mathbb{C}$, e $\alpha_i \neq 0$, $i = 1, 2$,*

$$\mathcal{K}_k(A, b) = \mathcal{K}_k(\alpha_1 A, \alpha_2 b).$$

3. Se $\alpha \in \mathbb{C}$,

$$\mathcal{K}_k(A, b) = \mathcal{K}_k(A - \alpha I, b).$$

4. Se W *é regular, então*

$$\mathcal{K}_k(W^{-1}AW, W^{-1}b) = W^{-1}\mathcal{K}_k(A, b).$$

Demonstração: Exercício 3

Observação 1.4. *Em relação ao teorema 1.3, o item 1 informa que os subespaços de Krylov são encaixados. O item 2 fala sobre a invariância por multiplicação por escalar e o item 3 sobre a invariância por translação. Finalmente, o item 4 mostra que uma transformação por similaridade na matriz, pode ser controlada, apesar de não gerar o mesmo espaço de Krylov.*

Uma outra caracterização de um subespaço de Krylov, $\mathcal{K}_k(A, b)$, pode ser feita se observamos que todo $v \in \mathcal{K}_k(A, b)$ pode ser escrito da forma

$$v = \alpha_0 b + \alpha_1 Ab + \ldots + \alpha_{k-1}A^{k-1}b.$$

Se definimos o polinômio $p(A)$ como

$$p(A) = \alpha_0 I + \alpha_1 A + \ldots + \alpha_{k-1}A^{k-1},$$

então, temos $v = p(A)b$. Por outro lado, qualquer vetor da forma $v = q(A)b$ onde $q(*)$ é um polinômio de grau menor do que k pertence a $\mathcal{K}_k(A, b)$ com isso podemos fazer uma caracterização polinomial dos subespaços de Krylov:

$$\mathcal{K}_k(A, b) = \{p(A)b;\ \ \text{grau}(p) < k\}.$$

1.2 Base para Subespaço de Krylov

Uma pergunta natural é sobre a conveniência do uso, em aritmética finita, de uma sequência de Krylov de vetores linearmente independentes, como base para o subespaço de Krylov gerado por eles. A seguinte

observação é apresentada em [120, pág. 298], aqui faremos uma pequena modificação (em itálico) para ser coerente com o restante do texto.

> Uma *matriz que contém uma* base para o espaço de Krylov da seguinte forma
>
> $$K_k = \begin{pmatrix} b & Ab & \dots & A^{k-1}b \end{pmatrix}$$
>
> não é adequada para uma implementação numérica. A razão é que com o aumento de k, as colunas de $\mathcal{K}_k$ passam a ser cada vez mais linearmente dependentes, já que elas tendem gradativamente a se aproximar do espaço gerado pelos *autovetores associados ao* ***autovalor dominante*** [4].

E completa, analisando o exemplo [120, pág. 266] a seguir:

Exemplo 1.1. *Um matriz diagonal A, de ordem 100, é gerada com autovalores*

$$1; 0,95; 0,95^2; 0,95^3; ...; 0,95^{99}.$$

Começando com um vetor qualquer u_1*, geram-se os vetores* $u_{k+1} = A^k u_1$.

O autor discute que apesar da convergência para o autovalor dominante ser lenta, a diminuição do ângulo[5] entre o subespaço gerado pelos vetores u_k e o autoespaço vinculado ao autovalor dominante é bem mais rápida (veja exercício 4 desse capítulo), e completa mostrando uma tabela da evolução do número de condicionamento da matriz cujas colunas são os vetores u_k:

k	condicionamento
5	5,8e+02
10	3,4e+06
15	2,7e+10
20	3,1e+14

[4] Autovalor dominante é aquele de maior módulo.
[5] Para definição de ângulo entre subespaços, consultar [44, pág. 256] ou [119, pág. 73].

Ou seja, as colunas são cada vez mais próximas da dependência linear[6]. E não basta tentar ortogonalizar a base do subespaço, pois como os vetores são quase linearmente dependentes o processo de ortogonalização seria numericamente instável. Uma solução para esse problema será discutida no capítulo 2, onde apresentaremos o método de Arnoldi.

Se, numericamente, ortogonalizar a matriz de Krylov é má ideia, em matemática exata não há maiores problemas, caso as colunas da matriz sejam linearmente independentes, e isso graças ao teorema a seguir.

Teorema 1.4 (Fatoração QR)**.** *Seja $A \in \mathbb{C}^{m\times p}$, onde posto de A é igual a* **p***. Então A pode se escrito de forma única*

$$A = QR,$$

onde $Q \in \mathbb{C}^{m\times p}$ tem suas colunas ortonormais e $R \in \mathbb{C}^{p\times p}$ é regular, triangular superior com elementos da diagonal principal positivos.

Demonstração: Exercício 5.

Esse resultado é visto nos cursos básicos de álgebra linear e o método apresentado para se construir essa fatoração é, em geral, o processo de Gram-Schmidt. No entanto, há outros, por exemplo: método de reflexões de Householder e método de rotações de Givens (ver 4.2.2 ou [54, cáp. 5]).

1.3 Métodos de Projeção

Sejam $\mathcal{K}_k$ e $\mathcal{L}_k$[7] dois subespaços de $\mathbb{C}^m$, ambos com dimensão k. Um método de projeção consiste[8] em, dado um valor inicial x_0, construir uma sequência $\left(x_k\right)$ de vetores de $\mathbb{C}^m$, que atendam as seguintes pro-

[6] Para uma análise detalhada desse fato para matrizes hermitianas ver [120, pág. 269].
[7] Neste seção, os espaços $\mathcal{K}_k$ e $\mathcal{L}_k$ não são necessariamente espaços de Krylov.
[8] Essa parte é baseada, principalmente, em [24, págs. 131-132] e em [101, cáp. 5].

priedades:

$$x_k - x_0 \in \mathcal{K}_k, \tag{1.3.1}$$
$$r_k = b - Ax_k \perp \mathcal{L}_k. \tag{1.3.2}$$

Como não se sabe *a priori* o valor exato da solução e a cada novo passo do método tem-se uma nova aproximação x_k, à primeira vista, o melhor que se pode fazer é calcular a diferença, $b - Ax_k$, chamada de **resíduo**. Como esses métodos têm origem em métodos de aproximação de funções (ver, por exemplo, [73, cáp. 4]), a condição (1.3.2) é usualmente denominada **condição de Petrov-Galerkin**.

As duas condições anteriores podem definir de forma única cada x_k, vejamos as condições necessárias para tanto. Sejam U_k e V_k matrizes em $\mathbb{C}^{m\times k}$, cujas colunas são bases para, respectivamente, $\mathcal{K}_k$ e $\mathcal{L}_k$. A condição (1.3.1) pode ser escrita como

$$x_k - x_0 = U_k a_k, \quad a_k \in \mathbb{C}^k,$$

logo, o resíduo da segunda condição torna-se

$$r_k = b - Ax_k = b - Ax_0 - AU_k a_k = r_0 - AU_k a_k.$$

Como o resíduo é ortogonal a $\mathcal{L}_k$, ele é ortogonal a todos os vetores desse subespaço, e temos

$$V_k^H r_k = 0 \Rightarrow V_k^H r_0 - V_k^H AU_k a_k = 0 \Rightarrow a_k = (V_k^H AU_k)^{-1} V_k^H r_0,$$

caso $V_k^H AU_k$ seja uma matriz regular. O teorema a seguir descreve duas condições suficientes para essa matriz ser regular.

Teorema 1.5. *Sejam U_k e V_k matrizes em $\mathbb{C}^{m\times k}$, cujas colunas são bases para, respectivamente, $\mathcal{K}_k$ e $\mathcal{L}_k$, então qualquer uma das duas condições a seguir garante a regularidade da matriz $V_k^H AU_k$:*

- *A é positivo-definida e $\mathcal{K}_k = \mathcal{L}_k$, ou*
- *A é regular e $\mathcal{L}_k = A\mathcal{K}_k$.*

Demonstração: Exercício 7.

Caso a matriz $V_k^H AU_k$ seja regular, podemos escrever

$$x_k = x_0 + U_k(V_k^H AU_k)^{-1}V_k^H r_0 \tag{1.3.3}$$

e

$$r_k = r_0 - AU_k a_k \Rightarrow r_k = r_0 - AU_k(V_k^H AU_k)^{-1}V_k^H r_0.$$

Ora, $P_k := AU_k(V_k^H AU_k)^{-1}V_k^H$ é uma matriz de projeção em [9] $A\mathcal{K}_k$ cujo núcleo é $\mathcal{L}_k^{\perp}$, assim como $(I - P_k)$ é uma projeção em $\mathcal{L}_k^{\perp}$ cujo núcleo é $A\mathcal{K}_k$. E a fórmula anterior pode ser escrita

$$r_k = (I - P_k)r_0.$$

Ilustrando, de forma matricial, a condição de ortogonalidade de r_k em relação ao subespaço $\mathcal{L}_k$.

Observação 1.5. *Na construção anterior, caso $\mathcal{K}_k = \mathcal{L}_k$, a condição de ortogonalidade* (1.3.2) *denomina-se* ***condição de Galerkin*** *ou de* ***Ritz-Galerkin****.*

Observação 1.6. *Explorando a equação* (1.3.3)*, temos*

$$\begin{aligned} Ax_k = Ax_0 + P_k r_0 \Rightarrow P_k Ax_k = P_k Ax_0 + (P_k)^2 r_0 \Rightarrow \\ \Rightarrow P_k Ax_k = P_k Ax_0 + P_k r_0 \Rightarrow P_k Ax_k = P_k b. \end{aligned}$$

Ou seja, x_k dá uma solução exata do problema $Ax = b$ quando restringimos o espaço de busca e o lado direito ao subespaço $A\mathcal{K}_k$.

Observação 1.7. *Com isso, pode-se ver que os métodos de projeção chegam à solução no máximo em m passos, sendo um método direto de solução, sem ser, no entanto, competitivo com o método de eliminação gaussiana. O fato é que se a projeção for feita em espaços adequados, os métodos de projeção podem ser utilizados como métodos iterativos interessantes.*

Os teoremas abaixo reúnem informações sobre algumas decorrências das condições suficientes de existência da matriz de projeção, enunciadas no teorema 1.5.

[9] Veja a seção A.6 do Apêndice A.

Teorema 1.6. *Se x é a solução de $Ax = b$. Se A é simétrica e positivo-definida e se $\mathcal{K}_k = \mathcal{L}_k$, então x_k minimiza o produto interno abaixo, o qual define uma norma vetorial,*

$$(A(x-y),(x-y))^{10}$$

no espaço afim $x_0 + \mathcal{K}_k = \{y;\ y = x_0 + z,\ z \in \mathcal{K}_k\}$.

Teorema 1.7. *Se A é regular e $A\mathcal{K}_k = \mathcal{L}_k$, então x_k minimiza a norma euclidiana*

$$\|b - Ay\|_2$$

no espaço afim $x_0 + \mathcal{K}_k = \{y;\ y = x_0 + z,\ z \in \mathcal{K}_k\}$.

Teorema 1.8. *Supondo que $V_k^H A U_k$ seja regular. Se existe um k tal que $A\mathcal{K}_k = \mathcal{K}_k$ e se b e r_0 pertencem a $\mathcal{K}_k$, então $r_k = 0$ e $x_k = x$, ou seja, chega-se à solução exata.*

Demonstração: Exercícios 8, 9 e 10.

Com os resultados apresentados até agora, estamos prontos para uma definição geral sobre métodos de projeção em subespaços de Krylov (**MPSK**). Como base usaremos a mesma notação, $\mathcal{K}_k$ e $\mathcal{L}_k$, para os espaços de projeção. Queremos resolver $Ax = b$. Partindo de um valor inicial x_0 e calculando o resíduo inicial, $r_0 = b - Ax_0$, $\mathcal{K}_k$ será o subespaço de Krylov $K_k(A, r_0)$. Ao variarmos $\mathcal{L}_k$ e ao usarmos diferentes tipos de projeção daremos origem a distintos MPSK.

Em particular apresentaremos no próximo capítulo o método da ortogonalização completa (FOM) e o de resíduo minimal generalizado (GMRES); no primeiro $\mathcal{L}_k = K_k(A, r_0)$ e no segundo $\mathcal{L}_k = AK_k(A, r_0)$.

Exercícios

1. Demonstre o teorema 1.1.

[10] Usamos a notação (u, v) para representar o **produto interno** entre os vetores u e v, a não ser que seja explicitamente dito, será sempre o produto interno canônico $\sum_{i=1}^{m} u_i \overline{v}_i$.

2. Detalhe a afirmação feita na demonstração do teorema 1.2 de que o coeficiente independente da variável, α_0, do polinômio mínimo de um vetor em relação a uma matriz regular é diferente de zero. Por que é necessária a hipótese de divisão dos polinômios?

3. Demonstre as propriedades enunciadas no teorema 1.3.

4. Em relação ao subespaços de Krylov do exemplo 1.1, faça um programa em Matlab, ou equivalente, para calcular a convergência dos vetores u_k em relação ao autovalor dominante da matriz, e faça um outro para calcular o ângulo entre o autovalor dominante e o subespaço gerado pelos vetores u_k. Trace os gráficos e observe a diferença entre esses processos de convergência.

5. Demonstre o teorema 1.4.

6. Em relação ao teorema 1.4, prove que $\mathrm{Im}(A(:, 1:l)) = \mathrm{Im}(Q(:, 1:l))$ para $p = 1:l$, ou seja, que as primeiras l colunas de A geram o mesmo subespaço vetorial que as primeiras l colunas de Q. Prove que $\mathrm{Im}(A) = \mathrm{Im}(Q)$.

7. Demonstre o teorema 1.5

8. Demonstre o teorema 1.6.

9. Demonstre o teorema 1.7.

10. Demonstre o teorema 1.8.

11. Em relação à segunda parte do teorema 1.5, podemos afirmar que cada base de $\mathcal{L}_k$ pode ser construída pela multiplicação dos vetores de uma base de $\mathcal{K}_k$ pela matriz A? E quanto a ortogonalidade dessas bases, o que podemos afirmar?

Capítulo 2

Métodos de Krylov Baseados no Procedimento de Arnoldi

Há uma unanimidade entre os pesquisadores da área de métodos iterativos para solução de sistemas lineares: não existe o melhor método para a solução de problemas com matrizes não-simétricas [86]. Outro ponto de vista comum é o de que, para matrizes não-normais, há muito ainda o que se trabalhar na compreensão dos fatores que influenciam na convergência dos métodos. Ainda outro consenso, é o da necessidade de precondicionadores para acelerar os métodos de Krylov, que estudaremos no capítulo 3. Na seção 2.1, o método de Arnoldi, um dos procedimentos seminais dos métodos de Krylov ao lado do método de Lanczos, é discutido a partir de uma motivação para a sua construção e várias de suas propriedades são relacionadas. Nas demais seções, discutimos dois métodos paradigmáticos: o FOM e o GMRES. A literatura sobre esses métodos é bastante vasta, estando consolidada, por exemplo, nos livros [24], [59], [101], [131]. Ao fim deste capítulo, tocamos levemente na questão de estabilidade do GMRES quando do uso do procedimento de Arnoldi baseado no método de reflexões de Householder ou no método de ortogonalização de Gram-Schmidt modificado, apresentando a bibliografia necessária ao estudo desse tema. Deixaremos para o capítulo 5, a apresentação de algumas das variantes do GMRES surgidas nos últimos 20 anos.

2.1 Método de Arnoldi

O nome de Arnoldi[1] aparece tanto ligado à solução de problemas de autovalores quanto à solução de sistemas lineares. Nesta seção, mostraremos o método de Arnoldi [5] para ortogonalizar uma base de um subespaço Krylov. Visando facilitar o desenvolvimento, vamos supor que o grau do polinômio mínimo de r_0 em relação a A é maior do que k.

Uma motivação interessante para o método de Arnoldi é apresentada em [80, pág. 337] e foi formulada originalmente por Kees Vuik. Queremos construir uma base ortonormal para $\mathcal{K}_k(A, r_0) = \langle r_0, Ar_0, \ldots, A^{k-1}r_0\rangle$, tal que $\mathcal{K}_k(A, r_0) = \langle v_1, v_2, \ldots, v_{k-1}, v_k\rangle$. Seja $V = \begin{pmatrix} v_1 & v_2 & \ldots & v_k \end{pmatrix}$, logo $V^H V = I$. Vale, também, observar que a matriz de Krylov associada a $\mathcal{K}_k(A, r_0)$, $K_k = \begin{pmatrix} r_0 & Ar_0 & \ldots & A^{k-1}r_0 \end{pmatrix}$, goza da seguinte propriedade:

$$\begin{aligned} AK_k &= \begin{pmatrix} Ar_0 & A^2r_0 & \ldots & A^k r_0 \end{pmatrix} = \\ &= \begin{pmatrix} Ar_0 & A^2r_0 & \ldots & A^{k-1}r_0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & \ldots & A^k r_0 \end{pmatrix} = \\ &= K_k \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & \ldots & \ldots & 1 & 0 \end{pmatrix} + A^k r_0 e_k^H. \end{aligned} \tag{2.1.1}$$

onde e_k é o k-ésimo vetor da base canônica. Como buscamos uma base ortonormal, o método usual é o da fatoração $K_k = QR$, onde Q é uma matriz $m \times k$, cujas colunas são vetores ortonormais, e R é uma matriz regular e triangular superior. Chamando de H_1 a matriz

[1] Walter Edwin Arnoldi (1917-1995) foi um engenheiro americano que publicou sua técnica em 1951, não muito distante do aparecimento do algoritmo de Lanczos. Arnoldi graduou-se em engenharia mecânica no Stevens Institute of Technology, Hoboken, New Jersey, em 1937 e o seu mestrado foi obtido na Harvard University em 1939. Durante sua carreira, trabalhou como engenheiro na Hamilton Standard Division da United Aircraft Corporation, aonde, com o passar do tempo, tornou-se pesquisador chefe da divisão. Aposentou-se em 1977. Apesar de sua pesquisa ter versado sobre propriedades mecânicas e aerodinâmicas de aeronaves e estruturas aeroespaciais, o nome de Arnoldi é mantido vivo graças ao seu procedimento de ortogonalização (traduzido pelos autores de [81]).

de Hessenberg que aparece em (2.1.1), teremos

$$AQR = QRH_1 + A^k r_0 e_k^H.$$

Mais algumas contas:

$$AQ = (QRH_1 + A^k r_0 e_k^H)R^{-1} \Rightarrow Q^H AQ = (RH_1 + Q^H A^k r_0 e_k^H)R^{-1} \Rightarrow$$
$$\Rightarrow Q^H AQ = R(H_1 + R^{-1}Q^H A^k r_0 e_k^H)R^{-1},$$

ora

$$H_2 := H_1 + R^{-1}Q^H A^k r_0 e_k^H = \begin{pmatrix} 0 & 0 & \dots & 0 & \vdots \\ 1 & 0 & \dots & 0 & \vdots \\ 0 & 1 & \ddots & \vdots & R^{-1}Q^H A^k r_0 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & \dots & \dots & 1 & \vdots \end{pmatrix},$$

é uma matriz de Hessenberg e RH_2R^{-1} também o será (ver exercício 1). E assim, vemos que a decomposição $Q^H AQ$ é uma matriz de Hessenberg superior. E, podemos tomar para V as colunas de Q.

Continuemos o raciocínio de Vuik, agora por nossa conta e risco. Vamos desenvolver a última coluna de RH_2R^{-1}. Ela será igual a $Q^H A^k r_0 R^{-1}(k,k)$, para isso bastando interpretar a multiplicação de matrizes como um produto externo. $R^{-1}(k,k)$ é igual a $1/\|\widetilde{Q}(:,k)\|_2$, onde $\widetilde{Q}(:,k)$ é o vetor $Q(:,k)$ antes da normalização, ou seja, $Q(:,k) = \widetilde{Q}(:,k)/\|\widetilde{Q}(:,k)\|_2$. E assim, comparando as últimas colunas das matrizes $Q^H AQ$ e RH_2R^{-1}, temos

$$Q^H AQ(:,k) = Q^H A^k r_0 / \|\widetilde{Q}(:,k)\|_2,$$

ou ainda

$$QQ^H A\widetilde{Q}(:,k) = QQ^H A^k r_0.$$

Ou seja, as projeções ortogonais de $A\widetilde{Q}(:,k)$ e de $A^k r_0$ no subespaço de Krylov $\mathcal{K}_k(A, r_0)$ são as mesmas. Não temos ainda a resposta final, mas uma boa pista, de como gerar os subespaço de Krylov sem

utilizar $A^k r_0$. Será que no processo de ortogonalização, ao substituirmos o vetor $A^k r_0$ pelo vetor $A\widetilde{Q}(:,k)$, estaremos gerando bases para o mesmo subespaço de Krylov $\mathcal{K}_k(A, r_0)$? Quais as condições sobre os vetores $A^k r_0$ e $A\widetilde{Q}(:,k)$ para que os subespaços gerados sejam os mesmos? A resposta positiva é o método de Arnoldi, algoritmos 1 e 2, e a justificativa está no exercício 2. O algoritmo 1 usa o processo

Algoritmo 1 Método de Arnoldi (A, r_0, k) - alternativa com Gram-Schmidt clássico

1: $V(:,1) = r_0/\|r_0\|$
2: para $j = 1 : k$
3: $\quad w = AV(:,j)$
4: $\quad H(1:j,j) = V(:,1:j)^H w$
5: $\quad w = (I - V(:,1:j)V(:,1:j)^H)w$
6: $\quad H(j+1,j) = \|w\|_2$
7: $\quad V(:,j+1) = w/H(j+1,j)$
8: fim-para

de ortogonalização de Gram-Schmidt, onde todos os escalares, que serão utilizados para multiplicar os elementos da base já existente, são calculados usando o mesmo valor de $w = AV(:,j)$. Esse procedimento é numericamente instável, e por razões de estabilidade uma versão modificada é utilizada [118][2], ver algoritmo 2.

Observação 2.1. *Em ambas as versões do algoritmo de Arnoldi ainda não há um teste sobre $H(j+1,j)$ ser numericamente zero (ou seja, menor que uma constante arbitrada). Esse fato, denominado de* ***ruptura*** *do algoritmo, ocorre quando o novo vetor pertence ao mesmo subespaço dos vetores gerados até aquele momento. Ou seja, quando $\mathcal{K}_k(A, r_0) \supset A\mathcal{K}_k(A, r_0)$ ou, ainda, $\mathcal{K}_k(A, r_0) = \mathcal{K}_{k+1}(A, r_0)$. Em uma real implementação computacional é necessária a inclusão de um teste de ruptura.*

[2]Para ver exemplo de instabilidade consultar, entre outros, [81, exemplo 5.5.5, pág. 316].

Algoritmo 2 Método de Arnoldi ($A,\ r_0,\ k$) - alternativa com Gram-Schmidt modificado

1: $V(:,1) = r_0/\|r_0\|_2$
2: para $j = 1 : k$
3: $\quad w = AV(:,j)$
4: $\quad$ para $i = 1 : j$
5: $\quad\quad H(i,j) = (V(:,i), w)$
6: $\quad\quad w = w - H(i,j)V(:,i)$
7: $\quad$ fim-para
8: $\quad H(j+1,j) = \|w\|_2$
9: $\quad V(:,j+1) = w/H(j+1,j)$
10: fim-para

Observação 2.2. *Em [119, pág. 279], o autor é bastante enfático quanto a inadequação do nome Gram-Schmidt modificado, uma vez que ele considera ser outro método com outras propriedades apesar da semelhança entre os algoritmos, para maiores detalhes ver a obra citada.*

Sejam $V_k \in \mathbb{C}^{m\times k}$, a matriz cujas colunas são os vetores $V(:,j)$, e $H_k \in \mathbb{C}^{k\times k}$, a matriz Hessenberg superior, formadas no procedimento de Arnoldi até a k-ésima iteração do algoritmo 2 antes de executarmos os passos 8: e 9:. O algoritmo completo, contadas as k iterações, terá uma representação matricial, até esse momento, dada por

$$AV_k - V_k H_k = we_k^H,$$

onde e_k é o k-ésimo vetor da base canônica. Ao incorporarmos os passos 8: e 9:, passamos a ter:

$$\begin{aligned} &AV_k - V_k H_k = H(k+1,k)V(:,k+1)e_k^H \Rightarrow \\ &\Rightarrow AV_k = V_k H_k + H(k+1,k)V(:,k+1)e_k^H. \end{aligned} \tag{2.1.2}$$

Temos aqui uma multiplicação entre matrizes representada por um produto externo e podemos escrevê-la de forma mais compacta como:

$$AV_k = V_{k+1}\overline{H}_k, \tag{2.1.3}$$

onde $V_{k+1} \in \mathbb{C}^{m\times(k+1)}$ e $\overline{H}_k \in \mathbb{C}^{(k+1)\times k}$. Há, ainda, uma relação simples a ser extraída:

$$V_k^H AV_k = H_k. \tag{2.1.4}$$

As fórmulas (2.1.2), (2.1.3) e (2.1.4) resumem algumas das propriedades do método de Arnoldi que usaremos adiante.

Observação 2.3. *Vale observar que a fórmula* (2.1.4) *nos lembra a decomposição de Schur (só que na decomposição de Schur, V_k é, necessariamente, quadrada). E, motivados por essa observação, nos próximos capítulos (ver seção 4.1 do 4 e 5.1 do capítulo 5) vamos utilizar autovalores relacionados à matriz H_k visando aumentar a velocidade de convergência dos métodos de Krylov baseados no procedimento de Arnoldi.*

2.2 Ortogonalização Completa - FOM

Para resolver $Ax = b$, o método da **ortogonalização completa** [98], [101] é um MPSK com as seguintes características: partindo de um valor inicial x_0, tem-se o resíduo inicial, $r_0 = b - Ax_0$. $\mathcal{K}_k$ será o subespaço de Krylov $\mathcal{K}_k(A, r_0)$. A cada nova iteração, calcula-se x_k impondo as condições: $(x_k - x_0) \in \mathcal{K}_k(A, r_0)$ e o resíduo $r_k = b - Ax_k$ deve ser ortogonal à $\mathcal{L}_k = \mathcal{K}_k(A, r_0)$. Nesse caso, o espaço de restrições será $\mathcal{L}_k = \mathcal{K}_k$ e $r_k \perp \mathcal{K}_k(A, r_0)$. Uma representação gráfica simplificada desse fato pode ser vista na figura 2.1.

Uma representação resumida da estrutura de uma iteração do FOM é apresentada no algoritmo 3. O primeiro passo do algoritmo 3 será feito pelo método de Arnoldi. No segundo passo, as condições dadas nos permitem detalhar as operações matriciais necessárias. Seja V_j uma base ortonormal para $\mathcal{K}_j(A, r_0)$, então temos que para algum $y_j \in \mathbb{C}^j$, $x_j - x_0 = V_j y_j$, o que atende à primeira condição. Quanto ao

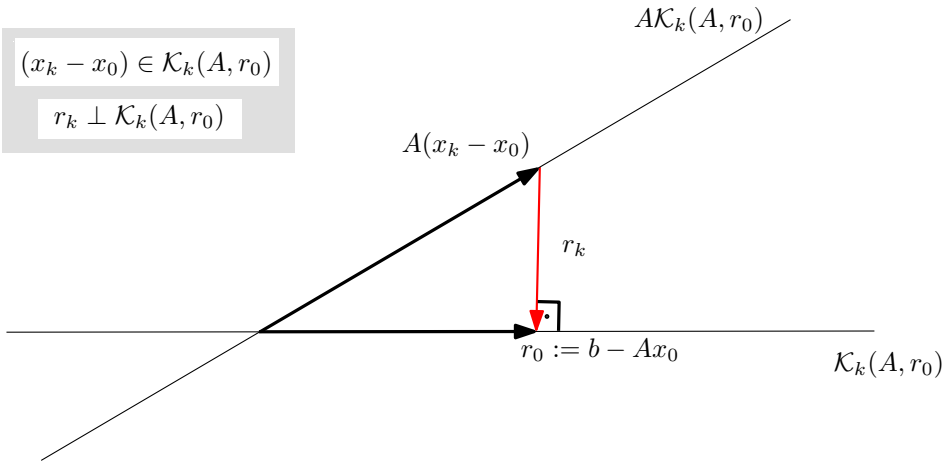

Figura 2.1: Representação esquemática da condição de ortogonalidade do resíduo do FOM.

Algoritmo 3 Ortogonalização completa $(A,\ x_0)$ - resumo de uma iteração

1: adicionar um vetor a uma base ortonormal para o subespaço de Krylov $\mathcal{K}_j(A, r_0)$,

2: calcular x_j tal que $x_j - x_0 \in \mathcal{K}_j(A, r_0)$ e que $r_j \perp \mathcal{K}_j(A, r_0)$.

resíduo, ele tem que ser ortogonal ao mesmo espaço, ou seja $r_j^H V_j = 0$ ou $V_j^H r_j = 0$, mas

$$r_j = b - Ax_j = b - A(x_0 + V_j y_j) = r_0 - AV_j y_j \Rightarrow V_j^H(r_0 - AV_j y_j) = 0 \Rightarrow$$
$$\Rightarrow V_j^H AV_j y_j = V_j^H r_0 \Rightarrow H_j y_j = V_j^H r_0.$$

Como essa base é ortonormal e o primeiro vetor da base é, exatamente, $r_0/\|r_0\|_2$, então $V_j^H r_0 = \big((r_0/\|r_0\|_2)^H r_0, 0, \ldots, 0\big)$. Logo, temos que resolver o sistema

$$H_j y_j = \|r_0\|_2 e_1,$$

onde e_1 é o primeiro vetor da base canônica de $\mathbb{C}^j$. Para que esse sistema tenha solução única é necessário e suficiente que H_j seja uma matriz regular. Essa condição não será sempre garantida e a singula-

ridade de H_j pode ocorrer em duas situações distintas. No primeiro caso, será uma ruptura benéfica do algoritmo:

$$H_j y_j = 0 \Rightarrow V_j^H A V_j y_j = 0 \Rightarrow V_j^H A \sum_{i=1}^{j} \alpha_i v_i = 0,$$

como as colunas de V_j geram uma base para $\mathcal{K}_j(A, r_0)$ temos ainda que

$$V_j^H A \sum_{i=1}^{j} \alpha_i v_i = 0 \Rightarrow V_j^H A \sum_{i=1}^{j} \alpha_i \left(\sum_{k=1}^{j} \beta_k A^{k-1} r_0 \right) = 0,$$

nesse caso, vamos considerar que $AV_j y_j = 0$

$$A \sum_{i=1}^{j} \alpha_i \left(\sum_{k=1}^{j} \beta_k A^{k-1} r_0 \right) = 0 \Rightarrow \sum_{i=1}^{j} \gamma_i A^i r_0 = 0,$$

ou seja chegamos ao polinômio mínimo de r_0 em relação a A e temos a solução exata.

Mas outra situação também pode ocorrer, nesse caso $z_j := AV_j y_j \neq 0$ e $V_j^H z_j = 0$, ou seja, existe um vetor não-nulo em $A\mathcal{K}_j(A, r_0)$ que é ortogonal a $\mathcal{K}_j(A, r_0)$, também nesse caso a matriz H_j será singular e haverá uma ruptura do FOM, sem ser benéfica (ver exercício 6).

O próximo resultado mostra como o cálculo do resíduo é simples para o FOM.

Teorema 2.1. *O resíduo da j-ésima iteração do FOM é dado por*

$$r_j = -\overline{H}_j(j+1, j) V_{j+1}(:, (j+1)) e_j^T y_j \quad e \quad \|r_j\|_2 = \overline{H}_j(j+1, j) |e_j^T y_j|.$$

Demonstração:

$$\begin{aligned} r_j &= b - Ax_j = b - Ax_0 - AV_j y_j = r_0 - V_{j+1}\overline{H}_j y_j = \\ &= r_0 - (V_j H_j + \overline{H}_j(j+1,j)V_{j+1}(:,(j+1))e_j^T)y_j = \\ &= \beta V_j e_1 - (V_j H_j + \overline{H}_j(j+1,j)V_{j+1}(:,(j+1))e_j^T)y_j = \\ &= V_j(\beta e_1 - H_j y_j) - \overline{H}_j(j+1,j)V_{j+1}(:,(j+1))e_j^T y_j = \\ &= -\overline{H}_j(j+1,j)V_{j+1}(:,(j+1))e_j^T y_j. \end{aligned}$$

Como $\overline{H}_j(j+1,j)e_j^T y_j$ é um escalar e $\|V_{j+1}(:,(j+1))\|_2 = 1$, temos os resultados. ■

2.3 Resíduo Minimal Generalizado - GMRES

O método de **resíduo minimal generalizado (GMRES)** [102] é um MPSK com as seguintes características. Para resolvermos $Ax = b$, partimos de um valor inicial x_0 e calculamos o resíduo inicial, $r_0 = b - Ax_0$. $\mathcal{K}_k$ será o subespaço de Krylov $\mathcal{K}_k(A, r_0)$, ou seja, $(x_k - x_0) \in \mathcal{K}_k(A, r_0)$, e o espaço de restrições será $\mathcal{L}_k = A\mathcal{K}_k(A, r_0)$ e, assim, o resíduo r_k é ortogonal a $A\mathcal{K}_k(A, r_0)$, $r_k \perp A\mathcal{K}_k(A, r_0)$. Com isso, o GMRES assegura que o resíduo, a cada iteração, não aumentará, no pior caso o resíduo das novas iterações será igual ao(s) da(s) anterior(es). Como a cada passo o espaço de busca está aumentando, mesmo depois de alguma **estagnação**, o método encontrará um ponto melhor. Uma representação gráfica simplificada desse fato pode ser vista na figura 2.2.

Algoritmo 4 GMRES (A, x_0) - resumo de uma iteração

1: adicionar um vetor a uma base ortonormal para o subespaço de Krylov $\mathcal{K}_j(A, r_0)$,

2: calcular x_j tal que $x_j - x_0 \in \mathcal{K}_j(A, r_0)$ e que $r_j \perp A\mathcal{K}_j(A, r_0)$.

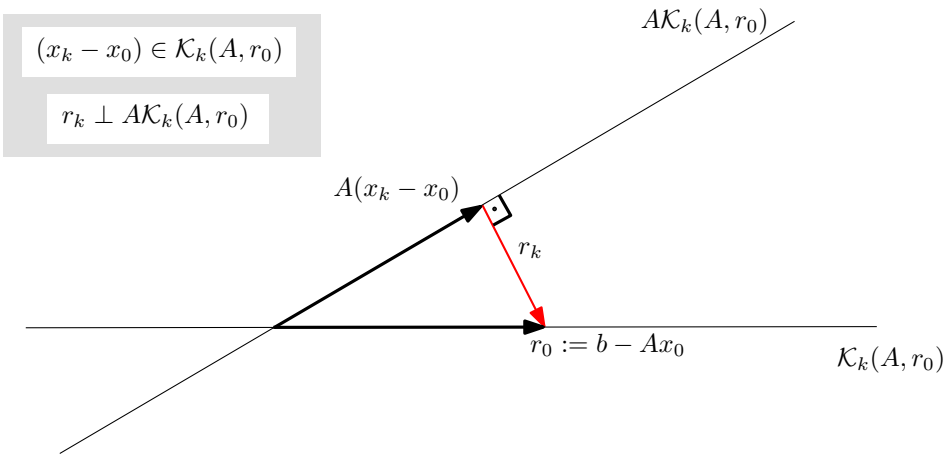

Figura 2.2: Representação esquemática da condição de ortogonalidade do resíduo do GMRES.

Uma representação resumida da estrutura de uma iteração do GMRES é apresentada no algoritmo 4, aonde o passo 1: será realizado através do método de Arnoldi e no passo 2: haverá a solução de um problema de quadrados mínimos através de uma fatoração QR adequada. Vejamos alguns dos detalhes desse processo:

$$r_k = b - Ax_k = b - A(x_0 + c_k) = r_0 - Ac_k, \quad c_k \in \mathcal{K}_k(A, r_0);$$

sejam V_{k+1} uma matriz cujas colunas formam uma base ortonormal para $\mathcal{K}_{k+1}(A, r_0)$ e V_k uma matriz cujas colunas formam uma base ortonormal para $\mathcal{K}_k(A, r_0)$, então $c_k = V_k y_k$, $y_k \in \mathbb{C}^k$, e podemos continuar o desenvolvimento acima, lembrando-nos de uma das relações de Arnoldi, $AV_k = V_{k+1}\overline{H}_k$:

$$r_k = r_0 - Ac_k = r_0 - AV_k y_k = r_0 - V_{k+1}\overline{H}_k y_k.$$

Na construção da base ortonormal consideramos $v_1 = r_0/\|r_0\|_2$, logo

$$r_0 - V_{k+1}\overline{H}_k y_k = \|r_0\|_2 v_1 - V_{k+1}\overline{H}_k y_k = \|r_0\|_2 V_{k+1} e_1 - V_{k+1}\overline{H}_k y_k$$

temos, então

$$r_k = V_{k+1}(\|r_0\|_2 e_1 - \overline{H}_k y_k),$$

mas $\|V_{k+1}\|_2 = 1$, uma vez que as suas colunas são vetores ortonormais, logo o problema de quadrados mínimos que temos que resolver é

$$\|r_k\|_2 = \min_{y_k \in \mathbb{C}^k} \| \|r_0\|_2 e_1 - \overline{H}_k y_k \|_2. \tag{2.3.5}$$

Desenvolvendo (2.3.5). Podemos construir o produto matricial

$$\| \begin{pmatrix} \overline{H}_k & \|r_0\|_2 e_1 \end{pmatrix} \begin{pmatrix} y_k \\ -1 \end{pmatrix} \|_2 . \tag{2.3.6}$$

Fazendo a fatoração QR de

$$\begin{pmatrix} \overline{H}_k & \|r_0\|_2 e_1 \end{pmatrix} = Q_{k+1} R_{k+1} = Q_{k+1} \begin{pmatrix} R_k & \zeta \\ 0 & s \end{pmatrix}.$$

E (2.3.6) pode ser escrita como:

$$\|Q_{k+1} \begin{pmatrix} R_k & \zeta \\ 0 & s \end{pmatrix} \begin{pmatrix} y_k \\ -1 \end{pmatrix} \|_2 = \| \begin{pmatrix} R_k & \zeta \\ 0 & s \end{pmatrix} \begin{pmatrix} y_k \\ -1 \end{pmatrix} \|_2 . \tag{2.3.7}$$

Logo a expressão (2.3.5), transforma-se em

$$\|r_k\|_2 = \min_{y_k \in \mathbb{C}^k} \| \begin{pmatrix} R_k y_k - \zeta \\ s \end{pmatrix} \|_2 = |s|. \tag{2.3.8}$$

O que fornece uma forma simples de se calcular a norma do resíduo, pelo menos em aritmética exata (ou infinita). Essa informação será útil tanto para detectar a convergência do método como para observar algum processo de estagnação[3].

Relembrando a observação 2.1, haverá um momento em que $H(k+1, k) = 0$, isso significa que o novo vetor calculado pertence ao espaço de Krylov anterior ou seja $w \in \mathcal{K}_k(A, r_0)$, confira o algoritmo 2. Ficará como o exercício 11 provar que essa **ruptura** do método é **benéfica**, pois chegou-se a solução do sistema linear.

A implementação do GMRES baseia-se na utilização de rotações de Givens para resolver o problema de quadrados mínimos (ver 4.2.2).

[3]No entanto um alerta deve ser feito aqui, pois esse cálculo, quando feito em aritmética finita, pode levar a erro, uma vez que a igualdade pode não estar garantida, para maiores esclarecimentos desse fenômeno consultar [32, pág. 90].

Como $\overline{H}_k$ é uma matriz de Hessenberg superior, as rotações são utilizadas para anular todos os valores que se encontram exatamente abaixo da diagonal principal. As rotações atuam apenas em uma entrada por vez, o trabalho feito anteriormente é aproveitado, sendo uma alternativa atraente por sua economia e estabilidade.

No artigo inicial sobre o GMRES [102] foram apresentados resultados de convergência do método para matrizes normais. No entanto, segundo [131], o principal resultado sobre a convergência do GMRES para uma matriz qualquer é, no mínimo, intrigante e foi apresentado em [60], em 1996.

Teorema 2.2 (Convergência do GMRES)**.** *Dada uma sequência não crescente de reais positivos $f_0 \geq f_1 \geq \ldots \geq f_{m+1}$ e um conjunto de complexos não nulos $\lambda_1, \lambda_1, \ldots, \lambda_m$, então existe uma matriz A com autovalores λ_j e com lado direito $b = f_0 e_1$ tal que os resíduos r_k do GMRES calculados na solução de $Ax = b$, com $x_0 = 0$, satisfazem $\|r_k\|_2 = f_k$, para $k = 0 : (n-1)$.*

Observação 2.4. *O teorema 2.2 nos informa que para uma matriz qualquer apenas os autovalores não são suficientes para caracterizar o comportamento da convergência do GMRES, ver exercício 12. No entanto, para matrizes normais, os autovalores são suficientes. Também para matrizes bem condicionadas, mesmo que não normais, os autovalores dão informação sobre o a convergência do método.*

Observação 2.5. *O outro lado da moeda do teorema 2.2 é que, na prática, ele não influencia o uso ou não do GMRES, apenas dá uma informação sobre casos possíveis e não sobre casos que sempre ocorrerão. Um outro aspecto é que o GMRES tem, em muitos casos importantes, uma convergência lenta, necessitando de precondicionadores para funcionar em um número de iterações aceitável, nesse caso a informação fornecida pelo teorema não tem grande aplicação.*

A discussão sobre as ferramentas matemáticas para caracterização da convergência do GMRES, e dos demais métodos de Krylov para matrizes não-normais, é uma área de estudo importante e que contém vários problemas em aberto, ver por exemplo [46], [90], [114], [141].

Um comentário necessário é sobre a estabilidade do método GMRES. Há dois resultados em [91] e [96] onde são caracterizadas a estabilidade em relação ao erro inverso das implementações do GMRES usando as reflexões de Householder e o método modificado de Gram-Schmidt no processo de Arnoldi. Esses resultados asseguram que pequenas modificações nos dados tratados não irão acarretar grande problemas à solução do problema, uma vez que se estará resolvendo exatamente um problema próximo. Ou seja a dificuldade será intrínseca ao próprio sistema que está sendo resolvido e não devida ao algoritmo utilizado. Trata-se de uma leitura técnica e importante para os pesquisadores da área.

A versão utilizada na prática para o GMRES é a **com recomeço**, ver por exemplo os códigos que estão disponíveis nas principais bibliotecas que implementam o GMRES (PETSc [10], Templates [11], MKL [69], Trilinos [105], Matlab [124], entre outras). Com o avanço do número de iterações do GMRES, o armazenamento dos vetores necessários e o tamanho dos problemas de quadrados mínimos a serem resolvidos começam a inviabilizar a aplicação do método. Há várias alternativas: a escolha de um subconjunto reduzido dos vetores já calculados (ver versões truncadas e com deflação no capitulo 5) e o recomeço após de um número fixado de iterações (versões com recomeço). A versão com recomeço padrão simplesmente testa a convergência depois de um número fixo de iterações e, caso não se tenha atingido a cota desejada, mantém-se apenas a última aproximação, descartando-se todos os demais vetores, e usa-se esse aproximação como valor inicial para calcular um novo resíduo inicial e começar uma nova aplicação do GMRES[4], ver análises em [83], [110] e [126]. A vantagem dessa alternativa é que como cada iteração garante o não aumento da norma euclidiana do resíduo, com o uso dessa solução, garante-se que estaremos partindo de um ponto, possivelmente melhor do que a primeira aproximação x_0, ver exercício 12. Apesar de drástica, essa alternativa é das mais usadas na prática. A bem da verdade, a alternativa com recomeço é um método de Krylov apenas durante cada ciclo do GMRES, uma vez que a cada recomeço um novo subespaço de Krylov é

[4]Cada ciclo completo de recomeço é denominado **ciclo** do GMRES.

construído, ou seja o método completo não fica dentro de um mesmo subespaço de Krylov que aumenta a cada ciclo completo.

Há dezenas de variantes do GMRES que foram desenvolvidas nos últimos 20 anos, num emaranhado de letras difícil de ser acompanhado mesmo pelos especialistas, ver por exemplo [103] e [114].

Como derradeiro comentário, sugerimos a leitura atenta do livro [101] de Y. Saad, um dos criadores do GMRES, nos diversos capítulos referentes ao GMRES, desde a sua formulação, passando pela convergência, discutindo implementações e precondicionadores, entre outros tópicos.

Exercícios

1. Prove que o produto de matrizes $R_1 H R_2$, onde R_1 e R_2 são matrizes triangulares superiores e H é uma matriz Hessenberg superior, tem como resultado uma matriz Hessenberg superior.

2. Prove que caso $v_1 = b/\|b\|_2$ e

$$\langle v_1, v_2, \ldots, v_{j-1}, v_j \rangle = \langle v_1, v_2, \ldots, v_{j-1}, Av_{j-1} \rangle,$$

para todo $j > 1$, então $\langle v_1, v_2, \ldots, v_j \rangle = \langle b, Ab, \ldots, A^{j-1}b \rangle$.

3. Dê exemplo de vetores que tem projeções ortogonais iguais em um subespaço qualquer, mas com projeções ortogonais não colineares no complemento ortogonal ao espaço dado.

4. Demonstre que os algoritmos Gram-Schmidt e Gram-Schmidt modificado geram os mesmos resultados.

5. Mostre que cada laço do processo de Arnoldi pode ser escrito como uma projeção ortogonal de um dado vetor em um dado espaço. Exibir os espaços, os vetores e as matrizes de projeção envolvidas nesse processo.

6. Dê exemplo de matriz e vetores que causem ruptura não-benéfica do FOM.

7. Recupere o código em Matlab do GMRES e o transforme no FOM. Procure a coleção Templates em http://www.netlib.org/-templates/index.html. Escreva um código que implemente ao mesmo tempo o FOM e o GMRES (a exceção de algumas linhas de teste).

8. Em matemática exata, o GMRES apresenta apenas rupturas benéficas, o que não é verdade para o FOM, dada a proximidade dos algoritmos, será possível continuar o método FOM após uma ruptura não-benéfica? Proponha uma alternativa.

9. Justifique a passagem da equação (2.3.7) para a equação (2.3.8).

10. Faça os detalhes do cálculo da equação (2.3.8).

11. Prove que no GMRES quando $H(k+1,k) = 0$, durante o procedimento de Arnoldi, significa que se encontrou a solução exata.

12. [131, exerc 6.11, pág 77] Sejam e_i os vetores da base canônica em $\mathbb{R}^m$. Seja A a matriz cujas as colunas são sucessivamente $e_2, e_3, \ldots, e_m, e_1$. Seja $b = e_1$ e comece o GMRES com $x_0 = 0$. Mostre que as matrizes de Hessenberg superiores associadas às bases ortonormais calculadas no processo de Arnoldi para os subespaços de Krylov com dimensão menores ou igual a m tem a parte triangular superior igual a 0. Use esse fato para mostrar que $\|r_j\|_2 = \|r_0\|_2$ para todo $j \leq m$. O que ocorre na $m-$ésima iteração? Quais são os autovalores da matriz A? Quais são os autovalores da matrizes de Hessenberg (valores de Ritz)?

Capítulo 3

Erros, Precondicionadores e Critérios de Parada

Dado um método iterativo devemos responder a, pelo menos, três perguntas. Qual a qualidade da solução conseguida? O método pode ser mais rápido? O método parou no momento correto, poderia ter parado antes, teria que parar depois? Esse capítulo tratará dessas questões. Na seção 3.1, vamos introduzir o conceito de erro inverso, usado tanto para a teoria sobre os métodos, quanto em aplicações práticas, na discussão sobre critérios de parada. Talvez um das áreas mais ativas nos métodos iterativos seja a da construção de precondicionadores para diminuir o custo total do método, assim como para garantir a confiabilidade do resultado. Apresentaremos um resumo sobre algumas alternativas na seção 3.2. Sobre os critérios de parada para os MPSK, apresentamos, na seção 3.3, uma discussão baseada no conhecimento atual da área, assim como um exemplo prático, a partir de um programa usado em aplicações industriais.

3.1 Erros e Qualidade de uma Solução

A principal referência dessa seção são as notas de curso de Serge Gratton [58], de 2008. Também foram consultadas as seguintes obras: [1], [32], [65], [121].

A análise de erro inverso foi introduzida por Givens e Wilkinson [138] e é um conceito poderoso para analisar a qualidade de soluções aproximadas:

1. é independente dos detalhes da propagação de erros de arredondamento. Os erros introduzidos durante os cálculos são interpretados como perturbações dos dados iniciais, e a solução calculada é considerada exata para o problema perturbado

2. já que os erros de arredondamento são vistos como perturbações nos dados, eles podem ser comparados a erros provenientes de aproximações numéricas ou medidas físicas.

Na realidade o erro inverso mede a distância entre os dados do problema inicial e os do problema perturbado, então ele depende dos dados que são permitidos variar e das normas utilizadas para medi-los. Para sistemas lineares há dois tipos de análise: uma baseada nas normas das matrizes e vetores envolvidos, e outra, baseada na medida de variação dos componentes individuais das matrizes e vetores, ver [32], [65]. Essas escolhas levam a fórmulas explícitas para o erro inverso que pode ser facilmente calculado. Para métodos iterativos, aconselha-se o uso do modelo de perturbação baseado nas normas das matrizes e vetores [3].

Iniciaremos por resultados sobre o erro direto relativo. Na solução de $Ax = b$, com A regular, $m \times m$, supomos que os dados do sistema, A e b, são submetidos a perturbações ΔA e Δb. A perturbação Δx resultante satisfaz a equação

$$(A + \Delta A)(x + \Delta x) = b + \Delta b. \tag{3.1.1}$$

Usando a norma euclidiana $\| * \|_2$, tanto vetorial quanto matricial, temos o seguinte resultado para o erro direto relativo:

Teorema 3.1. *Considerando a fórmula apresentada em* (3.1.1)*, em primeira ordem, temos a seguinte desigualdade:*

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq \|A\|_2 \|A^{-1}\|_2 \left(\frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta b\|_2}{\|b\|_2} \right). \tag{3.1.2}$$

Demonstração: Desenvolvendo (3.1.1)

$$Ax + \Delta Ax + A\Delta x + \Delta A\Delta x = b + \Delta b.$$

Descartando o termo de segunda ordem, $\Delta A\Delta x$, ficamos com $A\Delta x = \Delta b - \Delta Ax$, como consequência $\Delta x = A^{-1}(\Delta b - \Delta Ax)$. Como, por hipótese, A é regular, logo se $b \neq 0 \Rightarrow x \neq 0$, podemos então escrever

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq \|A^{-1}\|_2 \left(\|\Delta A\|_2 + \frac{\|\Delta b\|_2}{\|x\|_2} \right). \tag{3.1.3}$$

Como $b = Ax$ então $\|b\|_2 \leq \|A\|_2 \|x\|_2$, ou seja $\frac{1}{\|x\|_2} \leq \frac{\|\Delta A\|_2}{\|b\|_2}$, e podemos escrever

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq \|A\|_2 \|A^{-1}\|_2 \left(\frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta b\|_2}{\|b\|_2} \right),$$

uma vez que $AA^{-1} = I \Rightarrow \|A\|\|A^{-1}\| \geq 1$. ∎

Um segundo resultado sobre erro direto relativo na solução de problemas perturbados é dado pelo teorema 3.2, a seguir. Nesse caso, as hipóteses são modificadas e não se considera que há um desenvolvimento em primeira ordem, mas se impõe uma condição sobre o produto $\|\Delta A\|_2 \|A^{-1}\|_2$.

Teorema 3.2. *Considerando a fórmula apresentada em* (3.1.1)*, caso*

$$\|\Delta A\|_2 \|A^{-1}\|_2 \leq 1/2,$$

temos a seguinte desigualdade:

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq 2\|A\|_2 \|A^{-1}\|_2 \left(\frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta b\|_2}{\|b\|_2} \right). \tag{3.1.4}$$

Demonstração: Temos novamente

$$Ax + \Delta Ax + A\Delta x + \Delta A\Delta x = b + \Delta b.$$

Ficamos com $A\Delta x = \Delta b - \Delta A x - \Delta A \Delta x$, multiplicando por A^{-1}, temos $\Delta x = A^{-1}\Delta b - A^{-1}\Delta A x - A^{-1}\Delta A \Delta x$, utilizando as desigualdades entre normas e usando a hipótese fornecida, teremos

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq \|A^{-1}\|_2 \left(\|\Delta A\|_2 + \frac{\|\Delta b\|_2}{\|x\|_2} \right) + \frac{1}{2}\frac{\|\Delta x\|_2}{\|x\|_2}. \tag{3.1.5}$$

Considerando $\|b\|_2 \leq \|A\|_2 \|x\|_2$, chegamos na equação desejada. ■

O coeficiente $\kappa_2(A) := \|A\|_2 \|A^{-1}\|_2$ chama-se número de condicionamento da matriz A, e é o fator de amplificação das perturbações ΔA e Δb sobre os dados A e b, em norma relativa.

Suponhamos agora que temos uma aproximação $\widetilde{x}$ de x, obtida por exemplo (mas não necessariamente) por um cálculo em um computador. Chamamos de **erro inverso em relação a** A associado à $\widetilde{x}$, a quantidade

$$\eta_{(A)}(\widetilde{x}) = \min \big(\varepsilon > 0,\ \|\Delta A\|_2 \leq \varepsilon \|A\|_2, \quad (A + \Delta A)\widetilde{x} = b \big). \tag{3.1.6}$$

Por analogia, o erro relativo $\|\Delta x\|_2 / \|x\|_2 = \|\widetilde{x} - x\|_2 / \|x\|_2$, também pode ser chamado de **erro direto relativo**.

O erro inverso $\eta_{(A)}(\widetilde{x})$ mede, em norma induzida, a distância do problema exato A ao problema perturbado $A + \Delta A$ o qual $\widetilde{x}$ resolve exatamente. Esse erro inverso determina a medida relativa $\dfrac{\|\Delta A\|_2}{\|A\|_2}$ da perturbação de A que ocorre no cálculo da solução $\widetilde{x}$. Se $\widetilde{x}$ é o resultado de um cálculo em computador, o **cálculo** de $\widetilde{x}$ será **confiável** se o erro inverso associado é da ordem da precisão de máquina ϵ, ou seja

$$\eta_{(A)}(\widetilde{x}) \approx C\epsilon,$$

onde C é uma constante que pode depender dos dados do problema - aqui, A, b e m. Se, por acaso, a matrix A e/ou o lado direito b também carregam erros de cálculo, então o cálculo de $\widetilde{x}$ é confiável se o seu erro inverso associado, $\eta_{(A)}(\widetilde{x})$, é da ordem dos erros em A e/ou b.

Teorema 3.3. *Seja $r = b - A\widetilde{x}$ o vetor resíduo associado a $\widetilde{x}$. Então o erro inverso em relação à A é determinado pela fórmula*

$$\eta_{(A)}(\widetilde{x}) = \frac{\|r\|_2}{\|A\|_2\|\widetilde{x}\|_2}$$

Demonstração: Temos que

$$(A + \Delta A)\widetilde{x} = b \Rightarrow r = \Delta A\widetilde{x} \Rightarrow \|r\|_2 \leq \|\Delta A\|_2\|\widetilde{x}\|_2,$$

e assim

$$\theta := \frac{\|r\|_2}{\|A\|_2\|\widetilde{x}\|_2} \leq \frac{\|\Delta A\|_2}{\|A\|_2},\ \forall \Delta A.$$

Com isso temos que $\theta \leq \eta_{(A)}(\widetilde{x})$. Vamos mostrar que essa cota inferior é atingida, e, com isso, se transformará em um mínimo que atenderá a definição de erro inverso. Escolhamos uma perturbação particular δA de A da seguinte forma

$$\delta A = \frac{r\widetilde{x}^T}{\|\widetilde{x}\|_2^2}.$$

Primeiro verificamos que $(A+\delta A)\widetilde{x} = b$, ora a igualdade $A\widetilde{x} - \frac{r\widetilde{x}^T}{\|\widetilde{x}\|_2^2}\widetilde{x} = b$ é verdadeira. Continuamos

$$\|r\widetilde{x}^T\|_2 = \max_{\|y\|_2=1} \|r(\widetilde{x}^T y)\| = \max_{\|y\|_2=1} |\widetilde{x}^T y|\|r\|_2 = \|\widetilde{x}\|_2\|r\|_2, \qquad (3.1.7)$$

a última igualdade é válida graças à desigualdade de Cauchy-Schwarz. Com isso

$$\theta = \frac{\|r\|_2\|\widetilde{x}\|_2}{\|A\|_2\|\widetilde{x}\|_2^2} = \frac{\|r\delta A\|_2}{\|A\|_2}.$$

■

Agora vamos realizar perturbações tanto em A quanto em b e definir um outro erro inverso. Suponhamos que temos uma aproximação $\widetilde{x}$

de x. Chamamos de **erro inverso** associado à $\widetilde{x}$, a quantidade

$$\eta(\widetilde{x}) = \min \big(\varepsilon > 0,\ \|\Delta A\|_2 \leq \varepsilon \|A\|_2,\ \|\Delta b\|_2 \leq \varepsilon \|b\|_2, \\ (A + \Delta A)\widetilde{x} = (b + \Delta b)\big). \quad (3.1.8)$$

Com essa outra definição é possível demonstrar um teorema um pouco mais geral do que o teorema 3.3.

Teorema 3.4 (Teorema de Rigal e Gaches)**.** *Seja $r = A\widetilde{x} - b$ o vetor resíduo associado a $\widetilde{x}$. Então o erro inverso é determinado pela fórmula*

$$\eta(\widetilde{x}) = \frac{\|r\|_2}{\|A\|_2\|\widetilde{x}\|_2 + \|b\|_2}$$

Demonstração: Fazer exercício 5. ■

Observação 3.1. *A formulação do teorema 3.4 aqui apresentada está menos genérica do que a apresentada na obra original e do que em [48] ou [65]. Nesses trabalhos, no denominador da expressão usa-se uma matriz geral E no lugar de A e um vetor geral f no lugar de b. No entanto, achamos que essa formulação, no quadro de nosso trabalho, permite uma melhor compreensão da informação que o teorema contém.*

Como conclusão desta seção, e de uma maneira bem genérica, podemos afirmar que o erro dentro de um cálculo de $\widetilde{x}$ se exprime por:

$$\text{erro direto em } \widetilde{x} \quad \leq \quad \text{condicionamento} \times \text{erro inverso em } \widetilde{x}. \quad (3.1.9)$$

Ou seja, se o erro direto é grande, isso pode ser tanto devido ao problema a ser resolvido (número de condicionamento grande) e/ou quanto ao algoritmo usado (erro inverso grande). O papel do erro inverso é permitir a distinção, dentro do erro direto, entre a parcela que é devida ao problema em si e a parcela que é devida ao método utilizado (algoritmo) para resolver o problema. Essa afirmação tem uma série de decorrências matemáticas e para um maior aprofundamento desses tópicos recomendamos as seguintes leituras: [1, págs. 79-84], [32, págs. 71-75], [65, págs. 120-133] e [121].

3.2 Precondicionadores

As principais referências dessa seção serão [2, cáp. 10], [13] e [80, cáp. 8]. Utilizaremos, também, material extraído de [36], [78], [81], [117], [125], [127] e [128].

Michele Benzi, em [13], constata que nos últimos anos a classificação dos métodos de solução de sistemas lineares em diretos e iterativos é uma simplificação que não descreve com fidelidade o atual estágio de desenvolvimento da área. Observa que várias técnicas usadas na solução de problemas com matrizes esparsas por métodos diretos são correntemente usadas como precondicionadores em métodos iterativos, buscando tornar os métodos iterativos mais confiáveis. A segunda observação é que, enquanto os métodos diretos são principalmente baseados, ao menos para matrizes quadradas, em alguma versão da eliminação gaussiana, os métodos iterativos tem uma gama de opções vasta e diferenciada. Sendo assim, a classificação desses últimos como sendo pertencentes a uma só classe, pode causar mais confusão do que esclarecimento; apesar se não ser incorreta é incompleta.

Hoje em dia é reconhecido que o problema mais crítico no desenvolvimento de solvers eficientes é a construção de precondicionadores, e essa centralidade deverá se tornar cada vez mais evidente. Em apoio a essa afirmação vejamos o que pensam Trefethen e Bau [127, pág. 319]:

> Nada será mais central para a ciência da computação no próximo século[1] do que a arte de transformar um problema aparentemente intratável em outro cuja a solução pode ser aproximada rapidamente. Para métodos em subespaços de Krylov, isto significa precondicionamento.

Um bom precondicionador tem que acelerar a convergência do método, no mínimo, para compensar o custo de sua construção, mas o objetivo é sempre mais ambicioso. O difícil problema de se encontrar um precondicionador eficiente é que se deve identificar um operador M, linear ou não, que atenda a pelos menos, necessária mas não exclusivamente, às seguintes propriedades:

[1] Este livro é de 1997.

1. *De alguma forma M^{-1} é uma boa aproximação para A^{-1}.* Embora não haja uma teoria geral, pode-se dizer que M deve ser de tal forma que $M^{-1}A$, ou AM^{-1}, deve ser próxima da matriz identidade e que seus autovalores sejam aglomerados em uma pequena região do plano complexo, longe de 0;

2. *M seja eficiente.* De forma que o método precondicionado convirja muito mais rapidamente do que a não precondicionado, compensando largamente o custo de construção e armazenamento de M;

3. *M ou M^{-1} sejam construídas em paralelo.* Para explorar as várias arquiteturas atuais, que cada vez mais utilizam o processamento paralelo.

Algebricamente, precondicionar um sistema linear $Ax = b$ seria aplicar uma das três transformações:

Pela esquerda:

$$M^{-1}Ax = M^{-1}b. \tag{3.2.10}$$

Como o lado direito é alterado, é necessário se ter atenção no critério de parada utilizado, pois ele deve refletir o novo problema tratado.

Pela direita:

$$AM^{-1}y = b, \quad \text{com} \quad M^{-1}y = x. \tag{3.2.11}$$

Nesse caso o lado direito não é alterado.

Ambos os lados: Caso o problema tratado seja simétrico e positivo-definido, torna-se importante preservar essas propriedades, nesse caso ao se aplicar um precondicionador simétrico em ambos os lados garantem-se ambas as características[2]:

$$M^{-1/2}AM^{-1/2}y = M^{-1/2}b, \quad \text{com} \quad M^{-1/2}y = x. \tag{3.2.12}$$

[2]Para o método de Gradientes Conjugados precondicionado essa aplicação é implícita caso se faça um implantação correta, ver [108].

Para o GMRES, cabe lembrar, o precondicionador é usado durante o procedimento de Arnoldi quando da preparação do novo vetor, no passo 3: do algoritmo 2, pág. 31, ou seja, $w = Av_i$ passa a ser $AM^{-1}v_i = w$, para um precondicionamento pela direita, e $M^{-1}Av_i = M^{-1}w$ para um pela esquerda. Pela direita, teremos que resolver o novo sistema $Mz_i = v_i$ para depois realizar a multiplicação $w = Az_i$. De forma semelhante temos o procedimento pela esquerda. Vale lembrar, que apesar de usarmos a notação matricial, o operador M, pode inclusive ser não-linear, como veremos no capítulo 5, onde apresentaremos precondicionadores flexíveis na seção 5.3. Vamos descrever sucintamente algumas classes de precondicionadores.

3.2.1 Partições Clássicas

Há várias formas interessantes de se particionar uma matriz $A = M - N$. Uma boa escolha pode ser usada para acelerar uma iteração de ponto fixo do tipo

$$x_{n+1} = M^{-1}Nx_n + M^{-1}b, \tag{3.2.13}$$

que pode ser escrita

$$x_{n+1} = Gx_n + \widehat{b} \tag{3.2.14}$$

onde

$$G := M^{-1}N = (I - M^{-1}A), \quad \widehat{b} := M^{-1}b. \tag{3.2.15}$$

(3.2.13) é a iteração de ponto fixo para o sistema $Ax = b$, utilizando o precondicionador M^{-1}. As diferentes escolhas de M, levam a vários métodos, ver [101, págs. 284-287]. A condição para convergência dessa iteração é que o raio espectral da matriz precondicionada seja menor do que 1. Esta condição garante que o sistema precondicionado seja regular.

3.2.2 Fatorações Incompletas

A eliminação gaussiana é o algoritmo mais usado para resolver sistemas lineares densos. Mesmo para sistemas esparsos de grande porte com alguma estrutura, esse é o método mais usado e mais indicado.

No entanto, para matrizes muito grandes, muito esparsas e sem estrutura aparente que possa ser aproveitada, a fatoração LU pode levar a um preenchimento tal do fator U que a fatoração torna-se inviável. Para tentar aproveitar a força da fatoração LU, para sistemas que são necessariamente resolvidos por métodos iterativos, propuseram-se as fatorações incompletas como precondicionadores. A ideia é que a fatoração aproximada $\widetilde{L}\widetilde{U}$, seja o mais próxima de A, o que nem sempre é factível. Há algumas classes em que é possível provar a existência de uma fatoração incompleta, mas não há resultado geral de existência, mesmo quando da existência da fatoração LU completa.

Apresentaremos três alternativas, mas há outras, e para maiores detalhes as referências [33] e [101] devem ser consultadas.

A ideia básica é realizar a fatoração incompleta baseada em algum critério que pare a fatoração antes que os fatores completos tenham sido calculados. Alguns dos critérios usados são:

1. **Sem preenchimento:** nesse caso os fatores aproximados $\tilde{L}$ e $\tilde{U}$ terão estrutura tal que $\tilde{L}\tilde{U}$ tenha a mesma esparsidade do que A. Também denominada preenchimento 0.

2. **Preenchimento controlado por posição:** permite-se que algumas posições anteriormente nulas em A venham a ser não nulas no produto $\tilde{L}\tilde{U}$. Esse controle se dá através da análise de um grafo de dependências da matriz A, ver [51].

3. **Preenchimento controlado por valor:** permite-se algum preenchimento, baseado em que o novo elemento do fator esteja acima de um teto pré-estabelecido.

3.2.3 Inversa Aproximada

Nesse caso, ao invés de se construir uma aproximação de A, se constrói uma aproximação da inversa de A, daí o nome do precondicionador. Uma forma de se construir essa aproximação é minimizar, aproximadamente, a norma de Frobenius da matriz $\|I - AM\|_F$, onde M é a inversa aproximada e precisa ter uma certa estrutura, através da

fórmula

$$\|AM - I\|_F^2 = \sum_{j=1}^{n} \|(AM - I)e_j\|_2^2 \tag{3.2.16}$$

cuja solução do problema de minimização pode ser feita separando em n problemas independentes de quadrados mínimos, a serem resolvidos de maneira aproximada,

$$\min_{m_k} \|Am_k - e_k\|,\ k = 1 : n. \tag{3.2.17}$$

Esta alternativa é proposta em [61] para ser implantada em paralelo, e permite a construção de uma inversa aproximada. Para outras abordagens ver [14, 15, 16, 17, 37, 38, 57, 142].

3.2.4 Decomposição de Domínio

Esse classe de precondicionadores é bem adaptada à computação paralela. É baseada na ideia simples de dividir o domínio de definição do problema em vários subdomínios, e resolver em cada subdomínio um subproblema e reunir a informação completa após. Essa ideia simples tem uma grande variedade de decorrências, e está fortemente ligada à solução de equações diferencias parciais em domínios com geometria complexas ou quando equações diferentes são utilizadas em regiões diferentes do domínio de um mesmo problema. Grosso modo, podem-se dividir as abordagens de decomposição de domínio em duas famílias:

- O primeiro grupo recebe o nome de métodos de Schwarz. O domínio é dividido em subdomínios com recobrimento e subproblemas locais são resolvidos em cada subdomínio. A solução de um subdomínio se transforma em uma condição de fronteira para os subdomínios vizinhos, pois o recobrimento permite essa possibilidade. Esse método foi proposto por H.-A. Schwarz em 1870 [106] para provar a existência de soluções de problemas de equações diferenciais definidas em domínios com geometrias complexas, que separados em regiões mais simples, onde se conheçam as soluções, permite a prova de existência de solução para a região completa.

- O segundo grupo usa subdomínios sem recobrimento. É possível, nesse caso, dividir as incógnitas do problema em dois grupos: as que estão na interface dos subdomínios e as que se encontram nos diversos interiores de cada subdomínio. De forma completamente algébrica pode-se calcular a matriz do complemento de Schur das incógnitas da interface em relação às demais incógnitas. O problema é resolvido para a interface e a solução serve de condição de fronteira para os problemas internos que podem ser resolvidos de forma independente [28], [29] e [30]. Esse métodos recebem várias denominações, entre elas métodos de subestruturação ou métodos do complemento de Schur.

A quantidade e diversidade de métodos que surgiram nos últimos 20 anos é notável. Para um visão atual dos métodos de decomposição de domínio, recomenda-se a leitura dos livros [78], [117] e [125].

3.2.5 Multigrid

Em vários problemas de computação científica, quando do uso de métodos iterativos, se identifica o seguinte fenômeno. Após algumas iterações, o erro se torna suave, mas não necessariamente menor. Um dos princípios básicos do método de multigrid (multimalhas) [128] é exatamente o de buscar a suavização do erro, essa parte do método faz uso de suavizadores. O outro princípio básico do método é o seguinte: a quantidade que é suave em uma determinada malha pode ser, sem grande perda, ou mesmo perda alguma, aproximada em uma malha mais grossa, com, por exemplo, o dobro de tamanho em cada célula. E assim, caso se tenha certeza que o erro tornou-se suave, após algumas iterações, pode-se aproximar o erro por um procedimento adequado em uma malha mais grossa, e assim, nesse segundo momento, a iteração torna-se bem mais barata.

Essa abordagem cria uma sequência de problemas auxiliares e pode ser construída através de procedimentos geométricos ou algébricos. Se o problema original é definido em uma malha que foi obtida através de vários passos de refinamento, pode-se usar uma hierarquia entre as malhas para se definir operadores de transferência entre malhas mais

finas e mais grossas, nesse caso estaremos tratando do denominado Método Multigrid Geométrico [137]. Se, no entanto, uma hierarquia não é definida, os operadores de transferência podem ser construídos algebricamente, a partir da matriz do sistema, essa abordagem chama-se Método Multigrid Algébrico [97]. Sobre esses esquemas, suas implantações e demais aspectos do método consultar [23], [64] , [97], [128] e [136].

3.3 Critérios de Parada

As principais referências dessa seção são [11, pág. 57-61, seção 4.2], [48, págs. 5-6], também faremos uso de: [32], [58] e [65].

É importante saber parar um MPSK. A primeira constatação necessária é que um critério de parada é dependente do problema que está sendo resolvido: da qualidade dos dados de entrada, da possibilidade de se calcular alguma norma da matriz, dos métodos em uso, etc. Esses são alguns dos fatores que irão definir qual o critério de parada para um dado problema com um dado algoritmo. O fato é que o erro relativo (ou direto), em geral, é caro para se calcular ou não está, simplesmente, disponível. A alternativa mais comum é se basear no resíduo e em algumas normas disponíveis, e calcular o erro inverso, que, como vimos na equação (3.1.9), é um dos fatores limitantes do erro direto relativo. Da teoria de erros de algoritmos é sabido que o melhor que se pode exigir de um método em precisão finita é que seu erro inverso seja da ordem da precisão de máquina, logo essa alternativa, quando disponível, é utilizada.

Em [11, pág. 57-61, seção 4.2] são apresentados cinco critérios de parada. Seja *tol* a tolerância definida pelo usuário:

1. O primeiro é baseado no erro inverso

$$\frac{\|r_i\|}{\|A\|\|\widetilde{x}_i\| + \|b\|} \leq tol.$$

Esse critério dá origem à seguinte cota para o erro direto

$$\|e_i\| \leq \|A^{-1}\|\|r_i\| \leq tol\|A^{-1}\|\big(\|A\|\|\widetilde{x}_i\| + \|b\|\big) \tag{3.3.18}$$

2. O segundo não faz uso da norma da matriz

$$\frac{\|r_i\|}{\|b\|} \leq tol.$$

Uma limitação desse método é que se $\|A\|\ \|\widetilde{x}\| \gg \|b\|$, o que só ocorrerá se A for muito mal condicionada (ou seja, que tenha um número de condicionamento elevado) e $\widetilde{x}$ estiver muito próximo do espaço nulo de A, então será difícil para qualquer método atender esse critério. A seguinte cota superior para o erro direto tem origem nesse critério:

$$\|e_i\| \leq \|A^{-1}\|\|r_i\| \leq tol\|A^{-1}\|\|b\| \tag{3.3.19}$$

3. O próximo critério faz uso da norma da inversa de A, mas não utiliza b

$$\frac{\|r_i\|\|A^{-1}\|}{\|\widetilde{x}_i\|} \leq tol.$$

Esse critério de parada garante que

$$\frac{\|e_i\|}{\|\widetilde{x}_i\|} \leq \frac{\|A^{-1}\|\|r_i\|}{\|\widetilde{x}_i\|} \leq tol. \tag{3.3.20}$$

4. O quarto critério usa o erro inverso baseado nos valores absolutos das coordenadas das matrizes e vetores envolvidos[3]

$$\frac{|r_i|_j}{\left(E|\widetilde{x}_i| + f\right)_j} \leq tol.$$

Aqui E é uma matriz definida pelo usuário com entradas não-negativas, f é um vetor definido pelo usuário com entradas não-negativas, e $|z|$ define o vetor cujas entradas são os valores absolutos do vetor z. Esse critério tem várias aplicações, e pode ser utilizado em vários problemas distintos, dada a liberdade de escolha dos parâmetros E e f.

[3]Para maiores explicações sobre esse tipo de erro inverso consultar [32] [65], mas essencialmente, neste caso não se trabalha com as normas dos vetores e matrizes mas sim com o tamanho dos elementos individualmente, sendo possível definir relações entre a análise do erro inverso baseado em normas e a baseada nos componentes individuais, ver [65, tab. 7.2, pág 130].

5. O quinto critério apresentado é bastante utilizado, mas é desaconselhado pelos autores, e se trata de

$$\frac{\|r_i\|}{\|r_0\|} \leq tol.$$

A desvantagem desse critério é a sua forte dependência na solução inicial x_0. Caso $x_0 = 0$, esse critério é equivalente ao critério 2 e apresenta os inconvenientes já assinalados, caso x_0 seja muito grande e muito inacurado então $\|r_0\|$ será grande e a iteração pode parar antes do que deve.

Apresentamos um exemplo prático de critérios de parada implantados em um código que implementa o GMRES e tem sido usado em aplicações científicas [48].

Vamos mostrar apenas os parâmetros padronizados do pacote, nesse caso, para um problema sem precondicionamento é utilizada uma combinação de dois critérios, ambos baseados na razão $\frac{\|r_i\|}{\|b\|} \leq tol$. No primeiro, a norma do resíduo que se encontra no denominador, baseia-se no cálculo que está implícito no GMRES, como vimos na página 37, no entanto devido a observação apresentada em nota de rodapé na página 37, utilizar apenas esse critério é temerário. Sendo assim, após o GMRES ter convergido usando essa aproximação do resíduo, calcula-se o resíduo usando a fórmula usual, $b - Ax_i$, a norma desse resíduo passa a fazer parte do denominador e executam-se iterações com esse novo critério. O objetivo dessa escolha é diminuir o número de produtos matriz-vetor, que são um dos núcleos de cálculos mais caros de um MPSK (o outro é a aplicação do precondicionador).

No caso de métodos precondicionados, o primeiro critério será

$$\frac{\text{aproximação do resíduo}}{\|M_1^{-1}b\|} \leq tol. \tag{3.3.21}$$

e o segundo é

$$\frac{\|M_1^{-1}AM_2^{-1}z_j - M_1^{-1}b\|_2}{\|M_1^{-1}b\|} \leq tol, \tag{3.3.22}$$

onde $x_j = M_2^{-1} z_j$. É o mesmo procedimento que o anterior, primeiro uma convergência com o resíduo aproximado e depois com o calculado pela fórmula usual.

Exercícios

1. Faça os detalhes da equação (3.1.3).
2. Faça os detalhes da equação (3.1.5).
3. Faça os detalhes da demonstração em (3.1.7).
4. Mostre, assumindo a notação da seção 3.1 , que caso

$$\eta_{(A)}(\widetilde{x})\|A\|_2\|A^{-1}\|_2 \leq 1/2$$

   então

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq 2\|A\|_2\|A^{-1}\|_2\eta_{(A)}(\widetilde{x}).$$

5. Demonstre o teorema 3.4.
6. Uma outra definição para erro inverso é apresentada em [11, pág. 59]. Onde pode se ler:

   > o erro inverso relativo à norma é definido como o menor valor possível de
   >
   > $$\max\left\{\frac{\|\Delta A\|_2}{\|A\|_2}, \frac{\|\Delta b\|_2}{\|b\|_2}\right\}.$$

   Discuta a equivalência dessa definição com a apresentada na seção 3.1.
7. Demonstre a desigualdade 3.3.18.
8. Demonstre a desigualdade 3.3.19.
9. Demonstre a desigualdade 3.3.20.

Capítulo 4

Tópicos de Álgebra Linear

Serão apresentados nesse capítulo alguns dos conceitos necessários à compreensão de novos, mas também dos antigos, métodos de Krylov. As partes são independentes e devem ser estudadas, pelo menos, em duas situações: no caso de dificuldades na compreensão dos métodos discutidos nos próximos capítulos, ou por pura e saudável curiosidade matemática.

O primeiro tópico trata de aproximações de autovalores e autovetores que vêm surgindo na literatura da área nos últimos vinte anos. Registre-se que alguns desses conceitos fazem parte do arsenal de ferramentas dos matemáticos há muitas décadas. A seção 4.1 tratará desses tópicos. Serão demonstrados resultados que deverão ser úteis nos capítulos posteriores

Os métodos de solução para problemas de quadrados mínimos tem cerca de 200 anos de prolífica história, mas conturbada em seu início por disputas sobe a prioridade de seu descobrimento [50]. Apesar de ser objeto de vasta literatura, achamos conveniente tê-lo à disposição quando do estudo do GMRES, já que é um dos passos fundamentais desse método. Aproveitamos para apresentar, na seção 4.2, algumas ideias intuitivas sobre os métodos de solução desse problema e fornecemos sugestões de leituras complementares.

4.1 Pares de Ritz e Pares Harmônicos de Ritz

Discutiremos aproximações para autovalores das matrizes que aparecem durante a aplicação dos métodos iterativos para a solução dos sistemas lineares que estamos resolvendo. Há uma vasta literatura sobre o tema, por exemplo [12, 56, 84, 90, 116].

Vamos apresentar, inicialmente, uma abordagem feita em [115], onde os autores consideram que o melhor método para se ter uma boa aproximação de um autovalor de uma matriz simétrica e real em um dado subespaço é o método de Rayleigh-Ritz. Baseando-se em [93, seção 11.4], os autores dizem que esse método comporta-se bem para o cálculo de autovalores exteriores e seus autovetores associados, mas que no entanto, o mesmo não ocorre para autovalores interiores ao espectro da matriz [71], [82], [107]. Há estudos para se tentar ultrapassar os problemas com o cálculo dos autopares interiores e os autores citam os esforços feitos em [107] e, em particular, em [82], aonde a inversão do operador (no nosso caso da matriz) pode ser tratada implicitamente (veremos como, no teorema 4.4). O método proposto recebeu o nome de Rayleigh-Ritz harmônico em [90]. As aproximações, valores harmônicos de Ritz, correspondentes a esse método, e que veremos ainda nessa seção, têm recebido considerável atenção dada a sua ligação com polinômios de métodos iterativos para sistemas lineares baseados no resíduo minimal.

Antes de começarmos os conceitos propriamente ditos, vamos introduzir três resultados clássicos que discutem autovalores pelo viés de problemas de otimização. Segundo C. Meyer em [81, pág. 651], se V é uma matriz $m \times k$, com $k < m$, com colunas ortonormais (por exemplo, no processo de Arnoldi surge uma matriz como essa, ver (2.1.4)) então $V^H AV = \mathrm{H}$ não é uma transformação de similaridade, logo seria errado concluir que os autovalores de A são iguais aos autovalores de H. Apesar disso, é comum que os autovalores de H sejam uma boa aproximação para os autovalores extremos de A, em particular quando A é hermitiana. Veremos uma boa motivação dessa possibilidade, no teorema seguinte, quando da caracterização dos autovalores extremos de matrizes hermitianas. Lembremo-nos que, por serem reais, os autovalores de uma matriz hermitiana são ordenáveis,

$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$.

Teorema 4.1 (Teorema de Rayleigh-Ritz)**.** *Sejam λ_1 o maior autovalor de $A \in \mathbb{C}^{m\times m}$, matriz hermitiana, e λ_m o menor. Então*

$$\lambda_{\max} = \lambda_1 = \max_{\|x\|_2=1} x^H Ax = \max_{x\neq 0} \frac{x^H Ax}{x^H x} \tag{4.1.1}$$

$$\lambda_{\min} = \lambda_m = \min_{\|x\|_2=1} x^H Ax = \min_{x\neq 0} \frac{x^H Ax}{x^H x}. \tag{4.1.2}$$

A demonstração deverá ser feita no exercício 1.

Observação 4.1. *Essa forma de definir autovalores é denominada* ***formulação variacional*** *e os quocientes que aparecem no teorema 4.1 recebem o nome de* ***quocientes de Rayleigh-Ritz*** *.*

Vamos apresentar um resultado que estende essa caracterização para todos os demais autovalores de uma matriz hermitiana.

Teorema 4.2 (Teorema de Courant-Fischer)**.** *Os autovalores $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$ de uma matriz hermitiana $A \in \mathbb{C}^{m\times m}$ são*

$$\lambda_i = \max_{\dim \mathcal{V}=i} \min_{\substack{x\in\mathcal{V} \\ \|x\|_2=1}} x^H Ax \quad e \quad \lambda_i = \min_{\dim \mathcal{V}=m-i+1} \max_{\substack{x\in\mathcal{V} \\ \|x\|_2=1}} x^H Ax.$$

A demonstração desse teorema clássico, que é baseada, assim como a do teorema de Rayleigh-Ritz, na decomposição espectral de uma matriz hermitiana, pode ser vista em [67, pág. 179] ou [81, pág. 550].

Observação 4.2. *No caso em que $i = m$ a formulação* $\max\min$ *reduz-se à apresentada em* (4.1.2)*, quando $i = 1$ a formulação* $\min\max$ *torna-se igual a* (4.1.1)*.*

O próximo teorema é uma aplicação do teorema de Courant-Fischer e fornece informações sobre autovalores de matrizes relacionadas por transformações unitárias ou ortogonais.

Teorema 4.3 (Teorema de Entrelaçamento [120, pág. 42])**.** *Sejam $A \in \mathbb{C}^{m\times m}$, hermitiana, com autovalores $\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \ldots \geq$*

$\lambda_m = \lambda_{\min}$ e $U \in \mathbb{C}^{m\times n}$ com colunas ortonormais. Temos $U^H AU$ e seus autovalores $\mu_{\max} = \mu_1 \geq \mu_2 \geq \ldots \geq \mu_n = \mu_{\min}$. Então

$$\lambda_i \geq \mu_i \geq \lambda_{m-n+1}, \quad i = 1 : n.$$

Esse resultado recebe o nome de teorema de entrelaçamento porque caso $n = m - 1$ então

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \ldots \lambda_{m-1} \geq \mu_{m-1} \geq \lambda_m,$$

ou seja, os autovalores da matrizes são entrelaçados pelos autovalores aproximados. Caso U seja uma matriz identidade de ordem menor do que m então $U^H AU$ é uma submatriz principal de A, e esse resultado é válido também para submatrizes principais.

A seguir vamos caracterizar aproximações de autovalores e autovetores da matriz A, associada a um sistema linear. As caracterizações seguintes servem para matrizes quaisquer e não apenas para matrizes hermitianas, como os teoremas anteriores.

Definição 4.1 (Par de Ritz)**.** *Para qualquer subespaço $\mathcal{S} \subset \mathbb{C}^m$, um vetor $x \in \mathcal{S}$, com $x \neq 0$, é um **vetor de Ritz** da matriz $A \in \mathbb{C}^{m\times m}$ associado ao **valor de Ritz** $\theta \in \mathbb{C}$, se*

$$w^H(Ax - \theta x) = 0, \forall w \in \mathcal{S} \quad ou \quad Ax - \theta x \perp \mathcal{S} \tag{4.1.3}$$

*$(x, \theta) \in \mathcal{S} \times \mathbb{C}$ é chamado **par de Ritz**.*

Uma representação gráfica de um par de Ritz pode ser vista na figura 4.1. Ao examinar essa figura, e as próximas referentes a autovalores aproximados, devemos ter o cuidado de vê-las apenas como representações esquemáticas, uma vez que os valores de Ritz, da mesma forma que os autovalores, podem ser números complexos e, nesse caso, essa representação não é válida.

Observação 4.3. *Usando a notação da definição 4.1, sejam $V \in \mathbb{C}^{m\times n}$ uma matriz cujas colunas são ortonormais ($V^H V = I \in \mathbb{C}^{n\times n}$) e $\mathcal{S} := \mathrm{Im}(V)$. Sejam $v \in \mathbb{C}^n$, $z \in \mathbb{C}^n$, $x = Vv$ e $w = Vz$. A equação* (4.1.3) *pode ser escrita:*

$$z^H(V^H AVv - \theta v) = 0, \forall z \in \mathbb{C}^n, \tag{4.1.4}$$

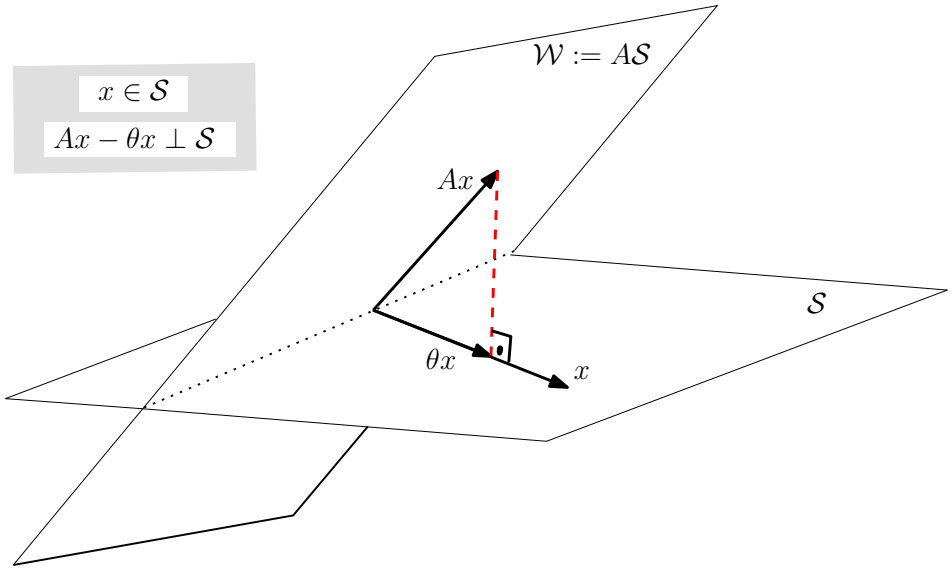

Figura 4.1: Representação esquemática de um par de Ritz.

que se torna um problema padrão de cálculo de autovalores

$$V^H AV\upsilon = \theta\upsilon, \tag{4.1.5}$$

onde $x = V\upsilon$*, com* $\upsilon \in \mathbb{C}^n$ *e* $x \in \mathbb{C}^m$*, ou seja,* υ *é a representação do vetor de Ritz* x *na base* $\{v_1, v_2, \ldots, v_n\}$*.*

A observação 4.3 nos permite fazer uma conexão entre a formulação variacional de Rayleigh-Ritz e o cálculo de um par de Ritz. Da equação (4.1.5) temos que

$$\upsilon^H V^H AV\upsilon = \theta\upsilon^H\upsilon \Rightarrow \theta = \frac{x^H Ax}{x^H x}.$$

Vale insistir que o teorema 4.1 tem como hipótese a matriz ser hermitiana e essa hipótese não é necessária à definição dos valores de Ritz.

Vamos a uma nova definição que será básica no desenvolvimento de algumas variantes do GMRES.

Definição 4.2 (Valor Harmônico de Ritz [90]). *Seja $\mathcal{S} \subset \mathbb{C}^m$. O escalar $\theta \in \mathbb{C}$ é um* ***valor harmônico de Ritz*** *de A em relação um dado espaço $\mathcal{W} \subset \mathbb{C}^m$, caso θ^{-1} seja um valor de Ritz de A^{-1} com relação $\mathcal{W} := A\mathcal{S}$.*

Uma representação gráfica para essa definição pode ser vista na figura 4.2. No entanto, usaremos uma outra formulação equivalente,

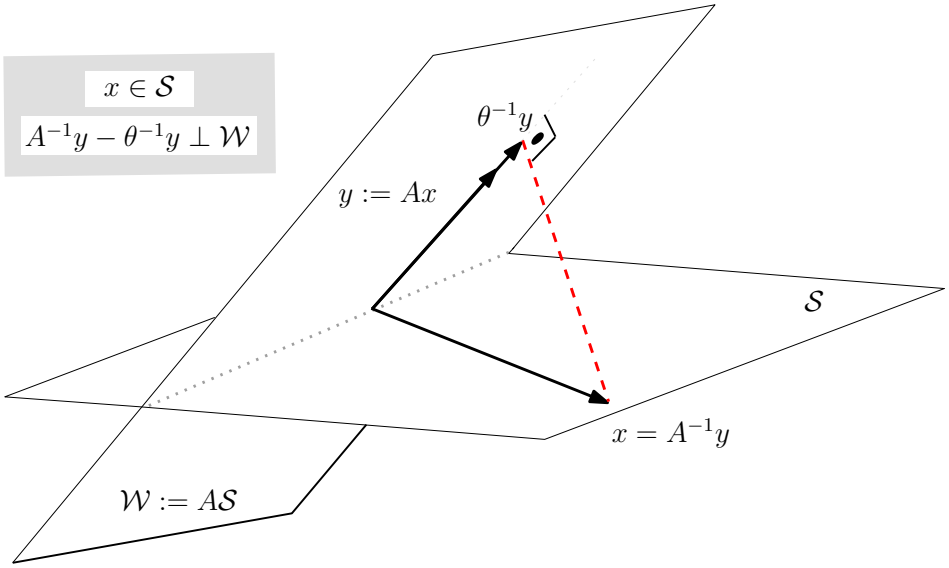

Figura 4.2: Representação esquemática de um par harmônico de Ritz usando a representação proposta na definição 4.2.

proposta em [116], para os valores harmônicos de Ritz.

Teorema 4.4 (Caracterização dos Pares Harmônicos de Ritz [116, Teorema 5.1, pág. 279]). *Sejam $\mathcal{S} \subset \mathbb{C}^m$ e $\mathcal{W} = \{y \in \mathbb{C}^m; \exists\ x \in S$ tal que $y = Ax\}$, ou seja, $\mathcal{W} := A\mathcal{S}$, então $\theta \in \mathbb{C}$ é um valor harmônico de Ritz de A em relação a $\mathcal{W}$, se e somente se,*

$$w^H(Ax - \theta x) = 0, \forall w \in \mathcal{W},\ \text{para algum } x \in \mathcal{S},\ x \neq 0. \qquad (4.1.6)$$

Denominaremos $x \in \mathcal{S}$ de ***vetor harmônico de Ritz*** *associado a θ, e $(x, \theta) \in S \times \mathbb{C}$ de* ***par harmônico de Ritz****. Uma representação gráfica para essa caracterização pode ser vista na figura 4.3.*

Demonstração: Pelas definições em 4.1 e 4.2, para θ ser um valor harmônico de Ritz de A em relação à $\mathcal{W}$, existem $y \neq 0 \in \mathcal{W}$ e $\theta \in \mathbb{C}$ tais que

$$w^H(A^{-1}y - \theta^{-1}y) = 0, \forall w \in \mathcal{W}, y \in \mathcal{W}, y \neq 0.$$

Basta apenas desenvolver para $y = Ax$, uma vez que $\mathcal{W} := A\mathcal{S}$

$$w^H(A^{-1}Ax - \theta^{-1}Ax) = 0 \Leftrightarrow \theta^{-1}w^H(\theta x - Ax) = 0 \Leftrightarrow w^H(\theta x - Ax) = 0.$$

■

É interessante observar que, no caso real, a equivalência entre as duas formulações de pares harmônicos de Ritz são simples relações de semelhança de triângulos retângulos, onde as hipotenusas são $x = A^{-1}y$, com $y \in \mathcal{W},\ y \neq 0$, quando usamos a definição 4.2, e θx, quando lançamos mão da caracterização proveniente do teorema 4.4.

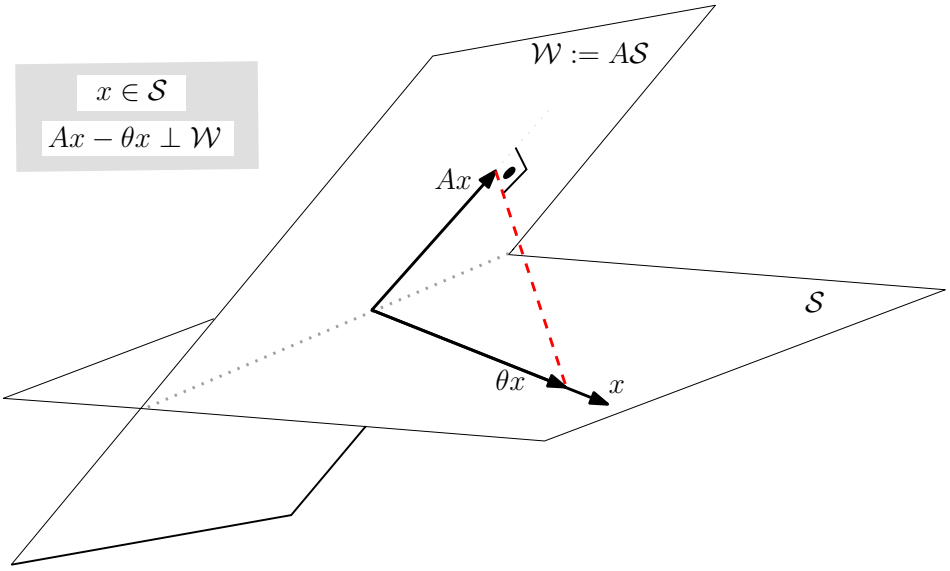

Figura 4.3: Representação esquemática de um par harmônico de Ritz usando a representação proposta no teorema 4.4.

Observação 4.4. *Usando a notação do teorema 4.4, sejam $V \in \mathbb{C}^{m\times n}$ uma matriz cujas as colunas são ortonormais, $V^HV = I \in \mathbb{C}^{n\times n}$ e $\mathcal{S} := \mathrm{Im}(V)$. Sejam $\chi \in \mathbb{C}^n$, $z \in \mathbb{C}^n$, $x = V\chi$ e $w = AVz$. A equação* (4.1.6) *pode ser escrita como:*

$$z^H(V^HA^HAV\chi - \theta V^HA^HV\chi) = 0, \forall z \in \mathbb{C}^n, \tag{4.1.7}$$

levando a um problema generalizado de autovalores*:*

$$V^HA^HAV\chi = \theta V^HA^HV\chi. \tag{4.1.8}$$

Caso V^HA^HV seja uma matriz regular, esse problema torna-se um problema de autovalores:

$$\left(V^HA^HV\right)^{-1} V^HA^HAV\chi = \theta\chi. \tag{4.1.9}$$

A observação 4.4 nos permite fazer a conexão entre uma variante da formulação variacional de Rayleigh-Ritz e o cálculo de um par harmônico de Ritz. Senão vejamos: da equação (4.1.8) temos que

$$\chi^HV^HA^HAV\chi = \theta\chi^HV^HA^HV\chi \Rightarrow \theta = \frac{(Ax)^HAx}{(Ax)^Hx}.$$

Aqui, novamente, vale o esclarecimento de que o teorema 4.1 tem como hipótese a matriz ser hermitiana; hipótese desnecessária à definição dos valores harmônicos de Ritz.

Os pares Ritz e os pares harmônicos Ritz, e algumas de suas variações [12], são bastante utilizados nos métodos iterativos para cálculo de autovalores, ver [8]. Para matrizes hermitianas e para matrizes que não estejam muito longe de serem normais, o comportamento dos autovalores e de suas aproximações ajudam a compreender o histórico da convergência de alguns métodos de Krylov [39], [60]. Com isso, os métodos de Krylov, quando aplicados à solução de um sistema linear, podem fazer uso de aproximações de autovalores, que estão implícitas, como veremos a seguir. A princípio, o GMRES com recomeço desconsidera a maior parte da informação guardada durante a iteração anterior, na seção 5.1, do capítulo 5, discutiremos como se pode

aproveitar a iteração anterior. Os próximos resultados fundamentam por que fazê-lo.

O teorema a seguir mostra a transformação do problema de cálculo de pares harmônicos de Ritz apresentado em (4.1.8) em um bem mais simples e demonstra uma propriedade relevante de ortogonalidade dos pares harmônicos de Ritz.

Teorema 4.5. *Seja V uma matriz cujas as colunas formam uma base ortonormal para $\mathcal{K}^{k+1}(A, r_0)$. Suponhamos que o polinômio mínimo de r_0 em relação a A tem grau maior que $k+1$. Usando a notação do método de Arnoldi, apresentado no algoritmo 2, na pág. 31, a equação para pares harmônicos de Ritz em* (4.1.8) *pode ser escrita como*

$$(H_k + h_{(k+1),k}^2 H_k^{-H} e_k e_k^T)\chi = \theta\chi. \tag{4.1.10}$$

Seja (x, θ) um par harmônico de Ritz de A em relação a $A\mathcal{K}_k(A, r_0)$, tal que $x = V_k\chi$, com $\chi \in \mathbb{C}^k$. Então vale a seguinte relação de ortogonalidade:

$$\overline{H}_k^H(\overline{H}_k\chi - \theta \begin{pmatrix}\chi\\0\end{pmatrix}) = 0. \tag{4.1.11}$$

Demonstração:

$$V_k^H A^H A V_k\chi = \theta V_k^H A^H V_k\chi$$

usando uma das relações provenientes do método de Arnoldi, $AV_k = V_{k+1}\overline{H}_k$, temos

$$\overline{H}_k^H V_{k+1}^H V_{k+1}\overline{H}_k\chi = \theta\overline{H}_k^H V_{k+1}^H V_k\chi$$

como V_{k+1} é ortogonal, podemos simplificar para

$$\overline{H}_k^H\overline{H}_k\chi = \theta\overline{H}_k^H \begin{pmatrix}I_{k\times k}\\0\end{pmatrix}\chi \Rightarrow \overline{H}_k^H\overline{H}_k\chi = \theta H_k^H\chi \tag{4.1.12}$$

escrevendo a matriz $\overline{H}_k$ em blocos, chegamos a

$$\begin{pmatrix} & & 0\\ H_k^H & & \vdots\\ & & 0\\ & & h_{(k+1),k}\end{pmatrix}\begin{pmatrix} & & & H_k\\ 0 & \dots & 0 & h_{(k+1),k}\end{pmatrix} = \theta H_k^H\chi$$

que pode ser simplificado para

$$(H_k^H H_k + h_{(k+1),k}^2 e_k e_k^T)\chi = \theta H_k^H \chi$$

como podemos assumir que H_k é regular, graças à hipótese sobre o grau do polinômio mínimo de r_0 em relação à A, então

$$(H_k + h_{(k+1),k}^2 H_k^{-H} e_k e_k^T)\chi = \theta\chi.$$

Provando a relação (4.1.10).

Partindo da equação (4.1.12)

$$\overline{H}_k^H \overline{H}_k \chi = \theta \overline{H}_k^H \begin{pmatrix} I_{k\times k} \\ 0 \end{pmatrix} \chi$$

com uma simples reorganização, temos

$$\overline{H}_k^H (\overline{H}_k \chi - \theta \begin{pmatrix} \chi \\ 0 \end{pmatrix}) = 0.$$

■

Cabe observar que o teorema anterior está fortemente ancorado no uso do método de Arnoldi para ortogonalização da matriz de Krylov.

O próximo teorema relaciona os resíduos dos cálculos dos valores harmônicos de Ritz com o resíduo de uma dada iteração do GMRES, ver [84].

Teorema 4.6. *Sejam $\overline{H}_k$ e $(\beta e_1 - \overline{H}_k y_k)$ provenientes do algoritmo do GMRES e sejam (x_i, θ_i) pares harmônicos de Ritz de A em relação a $A\mathcal{K}_k(A, r_0)$, tal que $x_i = V_k\chi_i$, com $\chi_i \in \mathbb{C}^k$. Suponhamos que o polinômio mínimo de r_0 em relação a A tem grau maior que $k+1$. Então*

$$\overline{H}_k \chi_i - \theta_i \begin{pmatrix} \chi_i \\ 0 \end{pmatrix} = \alpha_i(\beta e_1 - \overline{H}_k y_k), \tag{4.1.13}$$

com α_i escalares.

Demonstração: Pelo resultado encontrado em [101, corolário 1.39 do teorema 1.38, pág. 36], temos que $\beta e_1 - \overline{H}_k y_k$ é ortogonal a $\overline{H}_k y$, $\forall y \in \mathbb{C}^k$, onde

$$y_k = \arg\min_{y \in \mathbb{C}^k} \|\beta e_1 - \overline{H}_k y\|_2.$$

Então

$$(\beta e_1 - \overline{H}_k y_k) \perp \mathrm{Im}(\overline{H}_k).$$

Usando (4.1.11), podemos escrever que

$$(\overline{H}_k \chi_i - \theta_i \begin{pmatrix} \chi_i \\ 0 \end{pmatrix}) \perp \mathrm{Im}(\overline{H}_k).$$

Logo ambos os vetores pertencem ao $\mathrm{Nuc}(\overline{H}_k)$, mas pela hipótese sobre o grau do polinômio mínimo de r_0 em relação a A, H_k tem posto completo, $\mathrm{Nuc}(\overline{H}_k)$ tem dimensão 1, e

$$(\beta e_1 - \overline{H}_k y_k) \text{ é paralelo a } (\overline{H}_k \chi_i - \lambda_i \begin{pmatrix} \chi_i \\ 0 \end{pmatrix}).$$

■

Vamos utilizar o teorema anterior quando do estudo de variantes do GMRES nos capítulos 5 e 6.

4.2 Quadrados Mínimos

No GMRES um dos passos centrais é a solução de um problema de quadrados mínimos. Esse tipo de problema aparece em numerosas aplicações quando o resultado atingível é uma aproximação $\widetilde{x}$, tal que $A\widetilde{x}$ seja o mais próximo possível[1] de b. Em geral, essa abordagem ocorre quando o número de resultados experimentais é superior ao número de equações, levando a sistemas com matrizes retangulares. No caso de matrizes $m \times n$ onde $m > n$ pode ou não haver solução única. A formulação matemática[2] desse problema é

$$\|A\widetilde{x} - b\|_2 = \min_{x \in \mathbb{R}^m} \|Ax - b\|_2.$$

[1] Em relação a alguma medida aceitável, por exemplo a norma euclidiana do resíduo.
[2] Nessa parte da seção estamos tratando de problemas em $\mathbb{R}^m$.

Dizemos que $\widetilde{x}$ é a solução de um problema de minimização

$$\mathcal{P} : \min_{x \in \mathbb{R}^m} \|Ax - b\|_2. \tag{4.2.14}$$

4.2.1 Equações Normais

Vamos mostrar uma solução possível para o problema de minimização $\mathcal{P}$, (4.2.14). Trata-se de uma solução bastante usada em estatística. Em álgebra linear computacional, em geral, usa-se outra abordagem que veremos nas próximas seções

Teorema 4.7. *O problema $\mathcal{P}$ sempre admite ao menos uma solução. Uma condição necessária e suficiente para que $\widetilde{x}$ seja solução de $\mathcal{P}$ é que $\widetilde{x}$ seja solução da* ***equação normal***

$$A^T Ax = A^T b. \tag{4.2.15}$$

A solução $\widetilde{x}$ é única se e somente se A tem posto completo; neste caso $A^T A$ é positivo-definida.

Demonstração: Fazer exercício 3. ∎

Vamos agora apresentar uma das aplicações da decomposição de uma matriz em valores singulares. Seja $A \in \mathbb{R}^{m \times n}$, onde m e n são quaisquer. A **pseudo-inversa** de A é a matriz definida por $A^\dagger = V\Sigma^\dagger U^T$, aonde $U\Sigma V^T$ é a decomposição em valores singulares de A e $\Sigma^\dagger$ é a matriz transposta de Σ, com os coeficientes σ_i, os valores singulares de A, substituídos por seus inversos multiplicativos[3], $1/\sigma_i$.

Teorema 4.8. *Seja $\widetilde{x}$ a solução do problema de minimização $\mathcal{P}$ definido em* (4.2.14)*. Se A tem posto completo então*

$$A^\dagger = (A^T A)^{-1} A^T \quad e \quad \widetilde{x} = A^\dagger b.$$

Caso $m = n$, então $A^\dagger = A^{-1}$.

Demonstração: Fazer exercício 6. ∎

[3]Estamos considerando apenas o valores singulares positivos.

A descrição das equações normais sobre o corpo dos complexo poderá ser feita em separado para as partes real e imaginária, sem maiores dificuldades técnicas.

4.2.2 Solução por Fatoração QR

Vale observar que, na solução de sistemas lineares, as matrizes triangulares têm um papel destacado dada a facilidade de solução de sistemas triangulares. Se $A = QR$, com Q ortogonal ou unitária e R triangular superior, então $Ax = b \Rightarrow QRx = b \Rightarrow Rx = Q^H b \Rightarrow x = R^{-1}Q^H b$[4]. Esse é o método utilizado correntemente nos programas de solução de sistemas lineares por métodos de Krylov para resolver problemas de quadrados mínimos. Na verdade, trata-se de um outro método de solução direta de um sistema linear, só que computacionalmente mais caro do que a eliminação gaussiana, apesar de numericamente mais estável. Vamos apresentar três algoritmos que realizam a fatoração QR de uma dada matriz.

Reflexões de Householder

As reflexões de Householder[5] repousam sobre a ideia simples de ir transformando paulatinamente a matriz para a qual buscamos a fatoração QR, em uma matriz triangular superior R. Esse efeito será conseguido através de introdução de zeros na parte inferior de uma coluna por vez, um exemplo esquemático para uma matriz 4×3.

$$\begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \xrightarrow{Q_1} \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & * & * \\ 0 & * & * \end{pmatrix} \xrightarrow{Q_2} \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \\ 0 & 0 & * \end{pmatrix} \xrightarrow{Q_3} \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0 \end{pmatrix}$$

Por sinal, a mesma aparência da eliminação gaussiana, só que nesse caso as matrizes utilizadas para introduzir zeros nas colunas são matrizes ortogonais, ou seja: $Q_3 Q_2 Q_1 A = R$, ou então $A = Q_1^H Q_2^H Q_3^H R = QR$. As matrizes ortogonais Q_i, $i = 1 : 3$ são denominadas reflexões

[4] A solução, na prática, não se dá pela inversão de R mas pela solução do sistema triangular.
[5] Nessa parte estamos em $\mathbb{C}$.

ou refletores de Householder[6]. Alguns autores apresentam definições diferentes, apesar de equivalentes, para essa matrizes mas em todas elas a ideia subjacente é a mesma, através da soma ou subtração de duas projeções ortogonais construir uma reflexão, como veremos mais a frente. Seja o vetor $u_1 \in \mathbb{C}^k$, queremos anular todas as suas entradas, menos a primeira, preservando sua norma, logo, o novo vetor deverá ser $r_1 = (\|u_1\|, 0, \ldots, 0)$ ou $r_1 = (-\|u_1\|, 0, \ldots, 0)$ ambos em $\mathbb{C}^k$. No caso de vetores com coordenadas reais podemos observar a seguinte ocorrência: um fato da geometria plana básica é a perpendicularidade das diagonais de um losango, eis que esse fato aparece aqui. As direções das duas diagonais serão as direções em relação as quais se construirão as duas reflexões possíveis e os vetores u_1 e r_1 são os lados do losango. Esses fatos podem ser observados nas figuras de 4.4 a 4.7, onde mostramos a construção do fator R da fatoração QR da matriz $\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$ nesse caso estamos mostrando as duas reflexões possíveis, observa-se que os vetores da matriz ortogonal Q não estão representados.

A construção da matriz de reflexão de Householder pode ser vista como a subtração de duas projeções ortogonais. Seja w, o vetor unitário que servirá à construção da reflexão então estamos subtraindo duas projeções ortogonais $(I - ww^h)u_1$ e $(ww^h)u_1$ e construindo $(I - 2ww^h)u_1$ ou seu inverso aditivo, como apresentado nas figuras dessa seção. Outro fato representado nas figuras, são as retas α definidas pela condição de ortogonalidade ao vetor w Em dimensões maiores, os objetos α serão hiperplanos também ortogonais a w. Tendo apresentado essa motivação gráfica e geométrica passamos aos resultados teóricos.

[6] Alston Scott Householder (1904-1993) foi uma das primeiras pessoas a apreciar e promover o uso de refletores elementares para aplicações numéricas. Embora a sua dissertação de doutorado de 1937 na Universidade de Chicago ter sido sobre cálculo das variações, sua paixão era a biologia matemática, e essa foi a principal atividade de sua carreira até o momento em que esta foi interrompida pelo esforço de guerra, em 1944. Em 1946, Householder ingressou na Divisão de Matemática do Oak Ridge National Laboratory e se tornou seu diretor em 1948. Ele ficou em Oak Ridge durante o resto de sua carreira, e se tornou uma liderança mundial em análise numérica e no cálculo com matrizes. Como Givens (ver pág. 74), seu homólogo no Argonne National Laboratory, Householder foi um dos primeiros presidentes da SIAM.(traduzido pelos autores de [81])

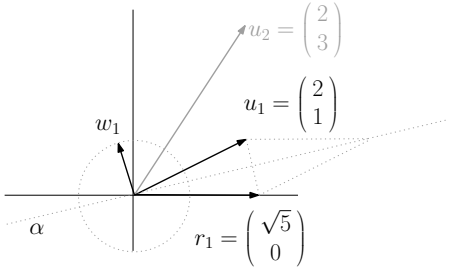

Figura 4.4: Cálculo do vetor unitário, w_1 com que se fará a reflexão.

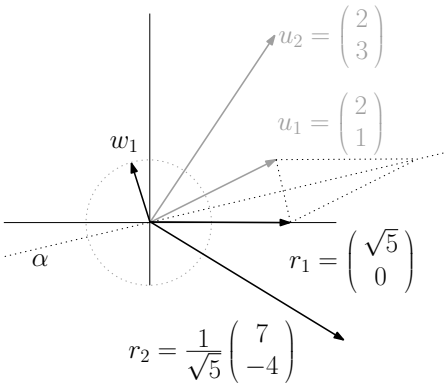

Figura 4.5: Os dois vetores já estão refletidos, formando as colunas de R, nesse caso 2×2.

Seja $w \neq 0$, $\|w\|_2 = 1$, a matriz $H(w) = I - 2ww^H$ é denominada de **matriz de reflexão de Householder** .

Teorema 4.9. *Seja $H = H(w)$ uma matriz de Householder com $\|w\|_2 = 1$. Seja $x \in \mathbb{C}^m$, tal que sua primeira componente, $x(1)$, seja real e maior que 0. Então $H(w)$ é uma matriz hermitiana e unitária, e o vetor*

$$w = \frac{x + \|x\|_2 e_1}{\|x + \|x\|_2 e_1\|} \tag{4.2.16}$$

define uma matriz de Householder $H := H(w)$ tal que $Hx = -\|x\|_2 e_1$.

Demonstração: Faça o exercício 7. ∎

Observação 4.5. *Se o vetor enunciado no teorema 4.9 não satisfizer a condição de ter sua primeira componente real e maior que 0, então, ainda assim, podemos definir matrizes de Householder:*

1. *caso $x(1)$ seja um complexo não nulo, então o vetor w deverá ser dado por*

$$w = \frac{x + \|x\|_2 \, \mathrm{e}^{\imath \, \arg(x(1))} \, e_1}{\|x + \|x\|_2 \, \mathrm{e}^{\imath \, \arg(x(1))} \, e_1\|} \tag{4.2.17}$$

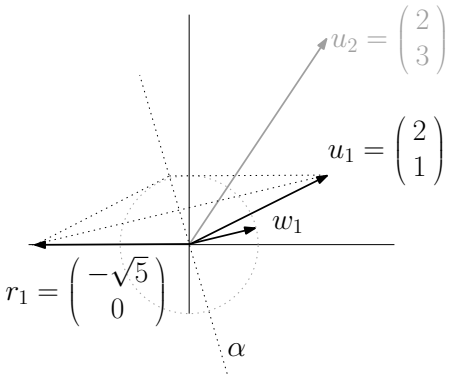

Figura 4.6: A outra reflexão possível, usando a outra diagonal do losango.

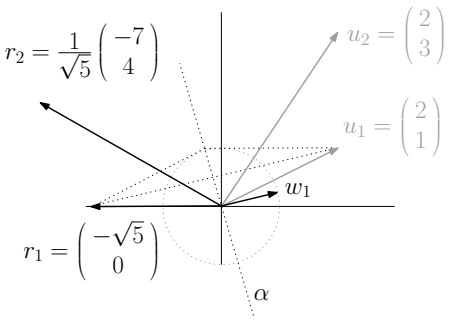

Figura 4.7: Os dois vetores já estão refletidos, usando a segunda reflexão possível, formando as colunas de R.

2. caso $x(1)$ real e negativo, então

$$w = \frac{x - \|x\|_2 e_1}{\|x - \|x\|_2 e_1\|} \tag{4.2.18}$$

Rotações de Givens

A próxima maneira de estabelecer uma fatoração QR é conhecida como método de rotações de Givens[7].

A ideia básica também é construir R através da transformação de todos os elementos abaixo da diagonal principal de A em elementos nulos. Nesse caso, cada elemento é anulado por vez. Assim como no caso de Householder, pode-se interpretar como a multiplicação pela

[7]James Wallace Givens, Jr. (1910-1993) foi um pioneiro no uso de rotações planas nos primórdios do cálculo automático de matrizes. Givens graduou-se no Lynchburg College, em 1928, e concluiu seu doutorado na Universidade de Princeton, em 1936. Depois de passar três anos no Instituto de Estudos Avançados, em Princeton, como assistente de O. Veblen, Givens aceitou sua nomeação na Cornell University, mais tarde transferiu-se para a Northwestern University. Além de sua carreira acadêmica, Givens foi Diretor da Divisão de Matemática Aplicada do Argonne National Laboratory, e, como Householder (ver pág. 72), seu homólogo no Oak Ridge National Laboratory, Givens foi um dos primeiros presidentes da SIAM (traduzido pelos autores de [81]).

esquerda por matrizes ortogonais. Seja a seguinte matriz

$$J(i,k,\theta) = \begin{pmatrix} 1 & 0 & \cdots & & & & & \\ 0 & 1 & \cdots & & 0 & & 0 & \\ \vdots & \vdots & \ddots & & & & & \\ & & & c & \cdots & s & & \\ & 0 & & \vdots & \ddots & \vdots & 0 & \\ & & & -s & \cdots & c & & \\ & & & & & & \ddots & \vdots & \vdots \\ & 0 & & & 0 & & \cdots & 1 & 0 \\ & & & & & & \cdots & 0 & 1 \end{pmatrix},$$

onde $c = \cos(\theta)$ e $s = \text{sen}(\theta)$. Ou seja, trata-se de uma matriz identidade onde quatro entradas são substituídas por senos e cossenos. O teorema a seguir ordena as ideias sobre as rotações de Givens.

Teorema 4.10. *Seja $A \in \mathbb{R}^{m\times n}$. Seja $k \neq i$, tal que $A(k,j) \neq 0$. Definindo θ tal que*

$$\cos(\theta) = \frac{A(i,j)}{\sqrt{(A(i,j))^2 + (A(k,j))^2}}$$

e

$$\text{sen}(\theta) = \frac{-A(k,j)}{\sqrt{(A(i,j))^2 + (A(k,j))^2}}$$

então o elemento de índice (k,j) do resultado do produto matricial $J(i,k,\theta)^T A$ é nulo.

Demonstração: Faça exercício 8. ■

Processo de Gram-Schmidt

Para a construção da fatoração QR de uma matriz A, o processo mais conhecido nos cursos básicos de álgebra linear é o **método de Gram-Schmidt**. Nesse caso podemos interpretá-lo como sendo a aplicação de transformações triangulares pela direita da matriz A, ver

[127]. Dado o amplo conhecimento desse método nos limitaremos a mostrar uma representação gráfica dele para uma matriz 2×2, ver figuras 4.8 a 4.11. Para o algoritmo dos métodos de Gram-Schmidt e o de Gram-Schmidt modificado ver a seção 2.1, aonde os métodos são apresentados para efetuar o procedimento de Arnoldi.

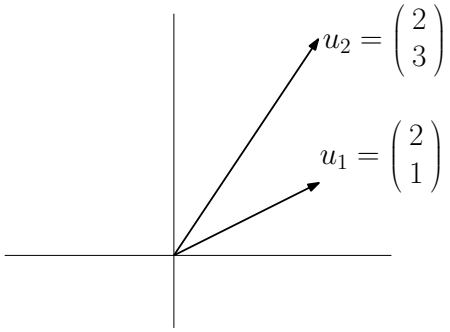

Figura 4.8: Vetores iniciais.

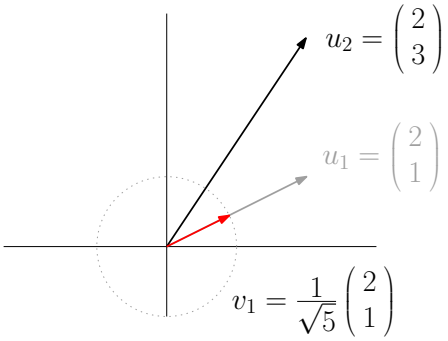

Figura 4.9: Normalização do primeiro vetor: $v_1 = u_1 / \|u_1\|_2$.

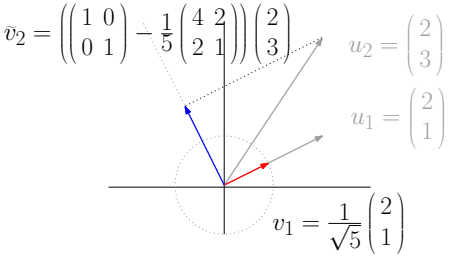

Figura 4.10: Projeção do segundo vetor: $\widetilde{v}_2 = (I - v_1 v_1^T) u_2$.

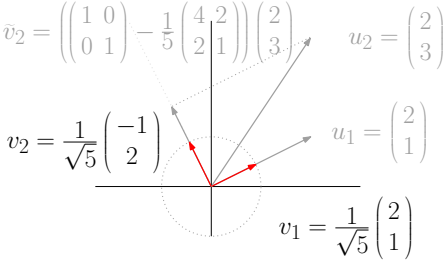

Figura 4.11: Normalização do segundo vetor: $v_2 = \widetilde{v}_2 / \|\widetilde{v}_2\|_2$.

Cabe observar que o processo de Gram-Schmidt modificado, não altera a construção do exemplo gráfico ora apresentado.

Comparação entre as fatorações QR

Para aclarar algumas das diferenças entre as quatro fatorações apresentadas, vamos apresentar um exemplo onde duas comparações serão

realizadas.

Seja $A \in \mathbb{R}^{25\times15}$, a matriz de Vandermonde formada a partir dos escalares $j\frac{1}{15}$, onde j é o índice da coluna. O condicionamento dessa matriz vale $\kappa_2(A) = 3,9 \times 10^9$. Sejam $\widetilde{Q}\widetilde{R}$ os fatores obtidos pelas quatro fatorações em um computador. Os erros residuais associados as fatorações QR são calculados usando a fórmula

$$\alpha = \frac{\|A - \widetilde{Q}\widetilde{R}\|_2}{\|A\|_2},$$

e seus valores podem ser vistos na tabela 4.1. Por esse critério, os procedimentos são equivalentes. Um outro valor relevante é a perda

Tabela 4.1: Comparação entre os erros das diversas fatorações QR.

Método	α
Gram-Schmidt	2×10^{-16}
Gram-Schmidt modificado	2×10^{-16}
Householder	10^{-15}
Givens	10^{-16}

da ortonormalidade, que pode ser medida por

$$\omega = \|\widetilde{Q}^H\widetilde{Q} - I\|_2.$$

Os diversos valores para ω podem ser vistos na tabela 4.2. E nesse

Tabela 4.2: Comparação da perda de ortogonalidade entre diversas fatorações QR.

Método	ω
Gram-Schmidt	$6,02$
Gram-Schmidt modificado	10^{-7}
Householder	2×10^{-15}
Givens	3×10^{-16}

caso os problemas e soluções aparecem, uma vez que Gram-Schmidt

não consegue construir uma base ortonormal, enquanto que a versão modificada consegue melhorar bastante; alguns autores chegam a afirmar que o nome Gram-Schmidt modificado não é apropriado, uma vez que se trata de outro método com outros resultados, apesar da semelhança entre os algoritmos. Os métodos de Householder e Givens fazem um bom trabalho segundo esse critério. Esses resultados podem ser consolidados teoricamente, e prova-se que as perdas de ortonormalidade desse métodos são proporcionais a $c_{mn}\,\kappa_2(A)^2\varepsilon$, para Gram-Schmidt, $c_{mn}\,\kappa_2(A)\varepsilon$, para Gram-Schmidt modificado, e de apenas $c_{mn}\,\varepsilon$ para Householder e Givens, onde ε é a precisão utilizada para executar as operações e c_{mn} é uma constante que cresce ligeiramente com o aumento da dimensão das matrizes.

4.2.3 Custo e Estabilidade

Quando da solução de problemas de quadrados mínimos, outro fator importante é o total de operações de ponto (na verdade, vírgula) flutuante (flops). É possível demonstrar que a solução usando equações normais custa $mn^2 + n^3/2$ flops, enquanto que as fatorações de Householder e Givens realizam $2mn^2 - 2n^3/3$. Ou seja, há uma diferença significativa entre as implementações, sendo a alternativa QR preferível apenas quando a estabilidade do método é essencial, ou não se tem informação suficiente sobre o número de condicionamento da matriz.

Outro aspecto importante é quanto à estabilidade dos vários métodos e nesse caso as alternativas de Householder e Givens são mais estáveis que o método das equações normais. Há uma extensa literatura sobre esse tema e os interessados podem consultar, por exemplo, [65] e [119].

Exercícios

1. Provar que o maior, λ_1, e menor, λ_m, autovalores de uma matriz hermitiana $m \times m$ podem ser ser descritos por

$$\lambda_1 = \max_{\|x\|_2=1} x^H A x \quad \text{e} \quad \lambda_m = \min_{\|x\|_2=1} x^H A x$$

2. Estabeleça uma relação entre os autovalores da matriz H_k que aparece no método de Arnoldi e os autovalores da matriz original A. Determine as restrições necessárias para as matrizes afim que as propriedades sejam válidas.

3. Faça a demonstração do teorema 4.7.

4. Faça um esboço gráfico da uma solução única para um problema de quadrados mínimos, relacionando $\mathrm{Im}(A)$ e $b - A\widetilde{x}$.

5. Seja $\mathbb{A}^{m\times n}$ com m e n quaisquer. Prove que a pseudo-inversa $A^{\dagger}$ de A é a única solução X das equações de Moore-Penrose[8]:

   (a) $XAX = X$,
   (b) $AXA = A$,
   (c) $(AX)^T = AX$,
   (d) $(XA)^T = XA$.

6. Faça a demonstração do teorema 4.8.

7. Faça a demonstração do teorema 4.9

8. Faça a demonstração do teorema 4.10.

9. Faça as implementações computacionais da fatoração QR:

   (a) usando reflexões de Householder.
   (b) usando rotações de Givens.

10. Faça uma implementação computacional dos algoritmos Gram-Schmidt e Gram-Schmidt modificado. Construa exemplos em que a instabilidade apareça (Sugestão: use três vetores quase linearmente dependentes, construídos com a constante *eps* do Matlab). Estude aonde exatamente surge a instabilidade .

11. Faça uma implementação computacional do procedimento de Arnoldi usando as reflexões de Householder para construção da base ortogonal.

[8] Ver definição de $A^{\dagger}$, também conhecida com inversa de Moore-Penrose, em [81, pág. 423].

12. Faça uma implementação computacional do procedimento de Arnoldi usando as rotações de Givens para construção da base ortogonal.

13. Provar que

$$(\mathbf{x}, \mathbf{y})_F := \text{traço } \mathbf{x}^H\mathbf{y}, \quad \|\mathbf{x}\|_F := \sqrt{\text{traço } \mathbf{x}^H\mathbf{x}}$$

são, respectivamente, um produto interno e uma norma bem definidos.

Capítulo 5

Novos desenvolvimentos dos Métodos de Krylov

O método GMRES completo[1] é, na maioria dos casos práticos, caro e apresenta uma convergência lenta. No capítulo 3, vimos que os precondicionadores são essenciais para aumentar a taxa de convergência do GMRES. Uma abordagem complementar ao desenvolvimento de novos precondicionadores é realizar modificações no próprio método. E assim, nos últimos 20 anos foram propostas várias alternativas de modificação do GMRES (assim como para os outros MPSK) para torná-lo mais adequado aos diferentes problemas das diversas áreas do conhecimento que fazem uso desse MPSK. Em [114], é apresentada uma exaustiva revisão sobre a evolução dos métodos de Krylov nas duas últimas décadas, e por isso trata-se de leitura obrigatória para qualquer interessado na área.

Neste capítulo, apresentamos alguns dos avanços ocorridos nesse período. Na seção 5.1, falaremos sobre métodos com recomeço baseados na recuperação de informação do subespaço de Krylov construído no ciclo anterior. Em particular, através da busca de aproximações de autovalores facilmente calculáveis usando a matriz de Hessenberg do ciclo anterior ([83], [84], [85]). Outro proposta, apresentada na seção

[1] Chamamos de GMRES completo aquele que só termina quando ocorre uma ruptura benéfica, ou quando se atinge uma redução do resíduo pré-estabelecida, guardando-se todos os vetores construídos.

5.2, é a preservação de subespaços do ciclo anterior mais relevantes para a garantia da ortogonalidade do método ([42], [45], [113]). Uma das propostas mais antigas de melhoria dos MPSK é apresentada na seção 5.3, tratam-se de métodos com precondicionadores que variam a cada iteração[2]. Em especial, precondicionadores que são eles próprios métodos de Krylov ([41], [99], [111], [112], [114], [132]). Um avanço mais recente que será apresentado na seção 5.4, está relacionado a se permitir que o produto matriz-vetor seja calculado de forma inexata, necessidade sentida em áreas aonde a matriz não é disponível ou é muito cara para ser calculada exatamente ([19], [20], [52], [112], [114], [129]). Apresentamos, a seguir, uma bibliografia onde as várias propostas aqui discutidas, e outras que não trataremos, podem ser estudadas: [9], [22], [25], [26], [34], [37], [49], [62], [63], [77], [89], [95], [100], [109] e [140].

5.1 Recomeço Deflacionado

Os artigos que fundamentam essa seção são os de Ronald B. Morgan [83, 84, 85] de 1995, 2000 e 2002, respectivamente, e o artigo de Luc Giraud e outros [53] de 2010.

Apesar do resultado teórico apresentado no teorema 2.2 (ver [60] e [86]), em um grande número de problemas importantes, a convergência dos MPSK depende em larga escala da distribuição dos autovalores. Quando da ocorrência de pequenos autovalores, a sua deflação (remoção), e dos autoespaços correspondentes, pode melhorar a taxa de convergência desses métodos [85]. Em [12], discutem-se algumas aproximações para os autopares de uma matriz e é provado um teorema sobre o entrelaçamento dessas aproximações com os autovetores de matrizes hermitianas [12, pág. 20]. Vamos discutir, nesta seção, as ideias relacionadas à deflação de autovalores aproximados, e seus autoespaços, quando do recomeço do GMRES. Para outros métodos de Krylov, para solução de sistemas lineares, também se pode utilizar técnica equivalente, ver [35, 104]. Essas alternativas também são

[2]Usaremos o termo **iteração do GMRES** para caracterizar a aplicação de um procedimento de Arnoldi e a solução de um problema de quadrados mínimos.

conhecidas como métodos de aumento por deflação [45].

5.1.1 GMRES-DR

A ideia central é o aumento dos subespaços de busca da solução aproximada com vetores harmônicos de Ritz herdados do ciclo anterior do GMRES[3]. Os principais trabalhos nessa direção são apresentados em [83, 84, 85] onde são propostos os métodos GMRES-E, GMRES-IR e GMRES-DR, respectivamente. Vamos resumir algumas das propostas e resultados formulados nesses artigos. As principais dificuldades de se utilizar a deflação em métodos com recomeço está ligada a dois fatos: como os espaços construídos são pequenos até a interrupção do ciclo, a deflação natural que poderia ocorrer para métodos sem recomeço (ver [87] e [133]) em geral, não acontece, em segundo lugar, guarda-se apenas um vetor em cada reinício, pelo menos nas versões tradicionais.

Vamos mostrar como construir espaços de busca com autovetores aproximados que ainda assim são subespaços de Krylov. Seja n a dimensão do subespaço construído até o momento de parada e seja k o número de autovetores aproximados utilizados em cada recomeço. Seja (y_i, θ_i) um par harmônico de Ritz. Temos a relação de Arnoldi surgida durante o processo de ortogonalização

$$AV_n = V_{n+1}\overline{H}_n, \tag{5.1.1}$$

onde $V_{n+1}^H V_{n+1} = I$, $\overline{H}_n \in \mathbb{C}^{(n+1)\times n}$ é uma matriz de Hessenberg e $H_n \in \mathbb{C}^{n\times n}$ é construída a partir de $\overline{H}_n$, com a exclusão de sua última linha.

O método usado pelo GMRES-E como espaço de busca para a solução aproximada é gerado por

$$\langle r_0, Ar_0, A^2 r_0, \ldots, A^{n-k-1} r_0, y_1, y_2, \ldots, y_k \rangle, \tag{5.1.2}$$

colocando ao final dos geradores, os vetores harmônicos de Ritz.

Pode parecer que colocando-se os vetores harmônicos de Ritz no início dos geradores, levaria a construção de um subespaço que não seja

[3]Usaremos o termo **ciclo do GMRES** para caracterizar o conjunto de iterações que ocorre ate à convergência ou até se atingir um número determinado de iterações, como ocorre na alternativa com recomeço. Em geral, um ciclo será formado por várias iterações.

de Krylov. No entanto, como provado em [84], os valores harmônicos de Ritz podem vir à frente. É demonstrado nesse artigo que o espaço gerado em (5.1.2) é equivalente a

$$\langle y_1, y_2, \ldots, y_k, Ay_i, A^2y_i, \ldots, A^{n-k}y_i, \rangle, \tag{5.1.3}$$

para $1 \leq i \leq k$, logo, ele contém todos os subespaços de Krylov que possuem vetores harmônicos de Ritz como vetores iniciais, ou seja, GMRES-E e GMRES-IR são equivalentes.

Em [139], os autores propuseram, tratando de cálculo de autovalores para problemas simétricos através de um método de Lanczos, uma forma particular de incorporação de valores de Ritz. Em [85], Morgan propõe a extensão dessa ideia para problemas não simétricos que usem o método de Arnoldi. A proposta desenvolvida nesse artigo é a seguinte: o primeiro ciclo do método é um GMRES com recomeço que produz um resíduo r_0. Seja V a matriz cujas colunas geram o espaço de busca em cada ciclo do método. Então, no começo do segundo ciclo, k valores harmônicos de Ritz são calculados e ortogonalizados através de um método QR. Esses vetores são colocados no início de V e r_0 é ortogonalizado em relação a $V(:, 1:k)$. Os próximos passos do método são o procedimento de Arnoldi para cálculo de $n-k$ vetores que formarão V_{n+1}, ver algoritmo 5. Pode ser provado que GMRES-E, GMRES-IR e GMRES-DR são matematicamente equivalentes.

Algoritmo 5 GMRES-DR - ciclos completos

1: Realizar um ciclo com n iterações de GMRES
2: faça até convergência
3: Calcular k valores harmônicos de Ritz, ortogonalizá-los e colocá-los em $V(:, 1:k)$ (ver exercício 1).
4: Ortogonalizar r_0 em relação a $V(:, 1:k)$.
5: Realizar o procedimento de Arnoldi para completar o subespaço e produzir V_{n+1} e $\overline{H}_n$.
6: Resolver o problema de quadrados mínimos, calcular a solução aproximada e o novo resíduo r_0.
7: fim-de-faça

Em [85], são demonstrados dois teoremas que provam que o subespaço gerado pelo método GMRES-DR é um subespaço de Krylov (uma demonstração simplificada desses resultados encontra-se em [45]). O primeiro resultado mostra que subespaço gerado pelos vetores harmônicos de Ritz e r_0 é de Krylov.

Teorema 5.1. *Seja o subespaço* $\mathcal{S} = \langle y_1, y_2, \ldots, y_k, v \rangle$ *tal que*

$$Ay_i - \theta_i y_i = \gamma_i v$$

para $\theta_1, \ldots, \theta_k$ *distintos e para* γ_i *não nulos. Então* $\mathcal{S}$ *é um subespaço de Krylov.*

Demonstração: Ver exercício 2. ■

O próximo teorema mostra que o subespaço completo é de Krylov.

Teorema 5.2. *O subespaço gerado por GMRES-DR é o subespaço* (5.1.2) *e é um subespaço de Krylov.*

Demonstração: Ver exercício 3. ■

Em [85], há algumas alusões sobre quais os tipos de valores harmônicos de Ritz - menores, maiores, randômicos - devem ser utilizados, mas nenhuma conclusão é estabelecida. No exercício 5, é sugerido o estudo dessa comparação.

5.2 Truncamento Otimal

Os artigos básicos desta seção são o de Eric de Sturler [42] de 1999, o de Michael Eierman e outros [45], de 2000, e o de Valeria Simoncini e Daniel B. Szyld [114], de 2007.

Os métodos da seção 5.1 baseiam-se na ideia de recomeçar um ciclo de um método de resíduo minimal, depois que o espaço de correção tenha chegado a uma dimensão n pré-estabelecida, compensando a perda de informação com o aumento do subespaço de busca através da deflação de autoespaços aproximados. Entretanto, como apontado

no teorema 2.2, e, em particular, para matrizes não normais e fortemente não simétricas, os autovalores aproximados podem dizer muito pouco sobre a convergência real do método. Além disso, a convergência lenta não é causada necessariamente por autovalores pequenos. Por exemplo, em [133] os autores mostram que quando alguns autovalores são equidistantes da origem, ou seja quando pertencem a um círculo cujo centro é a origem, o GMRES ficará estagnado por um número de passos igual ao número de autovalores que estão nesse círculo, independente do módulo dos autovalores. E mais, se os autovalores estão em um círculo de raio muito pequeno em torno da origem, pode até mesmo se tornar impossível o cálculo desses autovalores com acurácia aceitável.

No entanto, os métodos discutidos aqui estão relacionados aos da seção 5.1 pela tentativa de se reter alguma informação contida no espaço de busca do ciclo que terminou. Só que nesse caso, se tenta guardar a informação sobre ortogonalidade entre alguns subespaços específicos, considerados mais importantes para a convergência do método.

Ao invés de recomeçar apenas com um vetor, esses métodos truncam de alguma forma o espaço de busca. E, então, um subconjunto dos vetores da base construída até esse determinado momento é guardado. Em [42], De Sturler propões um esquema para descartar subespaços inteiros e não apenas vetores da base como em alguns outros métodos de truncamento (ver [101, págs. 172-177], QGMRES e DQGMRES). Esta seleção, proposta em [42], não se baseia nem em informações espectrais, nem na invariância de subespaços, mas em ângulos entre subespaços[4].

Vamos tratar de duas perguntas:

- Dada uma iteração $s < n$, onde n é a dimensão máxima permitida para o subespaço de busca. Se o método tivesse sido recomeçado após s iterações, como isso influenciaria a convergência? Ou seja, recomeça-se, fazem-se mais $(n - s)$ iterações, usando como valor inicial x_s e dispensando-se todos os demais vetores e matrizes já calculados.

[4]Para definição de ângulo entre subespaços, consultar [44, pág. 256] ou [119, pág. 73].

- Quais os subespaços das primeiras s iterações que deveriam ser preservados, afim de que nas próximas $(n-s)$ iterações se conseguisse uma convergência o mais próxima possível da convergência do GMRES completo, com n iterações?

A ideia do método é calcular os valores singulares de alguns dos subespaços em questão e descartar os que forem associados a pequenos valores singulares, uma vez que os vetores singulares associados não teriam maior contribuição para a convergência. O mesmo fato pode ser interpretado da seguinte forma. Dado o subespaço de Krylov $\mathcal{K}_s(A, r_0)$, com $s < n$, quais são as direções importantes para a convergência, no sentido de ao se manter a ortogonalidade em relação a essas direções, havendo o recomeço após s iterações, se conseguir a maior redução possível no resíduo. Os subespaços a serem comparados são $A\mathcal{K}_s(A, r_0)$ e $A\mathcal{K}_{(n-s)}(A, r_s)$. A comparação entre os subespaços nesse caso é barata uma vez que ambos estão em $\mathcal{K}_n(A, r_0)$. Logo os cálculos dos ângulos podem ser feitos em relação à base desse espaço e envolvem apenas matrizes pequenas.

A seleção de subespaços acima pode ser usada para melhorar o método GCRO[5] [41] de iterações internas-externas (ver seção 5.3). O método resultante se chama GCROT e transfere essa seleção de subespaços da iteração interna para a externa. Segundo [113], a versão atual do método precisa da seleção de seis parâmetros, que podem ser difíceis de serem ajustados, no entanto, ainda segundo [42], com alguma experiência, alguns dos parâmetros podem ser determinados facilmente.

5.3 Precondicionadores Flexíveis

O artigo base dessa seção é o de Valeria Simoncini e Daniel B. Szyld [111] de 2002, ao qual foram adicionadas algumas referências atualizadas e comentários dos seguintes artigos: [41], [53], [99], [132] e [134].

[5] O método GCRO é matematicamente equivalente ao GMRES pois constrói os mesmos espaços, tem as mesmas soluções intermediárias e resíduos parciais iguais, apesar de ser mais caro, apresenta maior flexibilidade na construção dos subespaços de Krylov.

A área de precondicionadores variáveis ou flexíveis vem apresentando um desenvolvimento constante nos últimos anos, ver as referências: [4], [6], [41], [53], [55], [88], [99], [111], [123], [132], [134], além das obras citadas nesses trabalhos. A ideia central é o uso de diferentes operadores em cada iteração (até mesmo operadores não lineares) como precondicionadores de um MPSK. Mesmo outros métodos de Krylov podem ser usados como precondicionadores, ver [41], [53], [99] e [132].

Dado o sistema linear $Ax = b$, aplicar um precondicionador padrão pela direita, como já vimos, consiste em substituir o sistema linear por

$$AM^{-1}y = b, \quad \text{com} \quad Mx = y \quad \text{ou} \quad M^{-1}y = x, \tag{5.3.4}$$

para um precondicionador adequado M. Uma das motivações para métodos com precondicionadores variáveis é se encontrar casos relevantes aonde se necessita resolver apenas aproximadamente

$$Mz = v,$$

considerando-se que M é, já, uma aproximação de A. Com isso poderá haver um M diferente para cada passo k do método de Krylov, pelo menos implicitamente. Uma outra possibilidade é a de se utilizar informações das iterações anteriores para melhorar a qualidade dos precondicionadores, ver [7], [47], [72].

Vamos analisar, nesta seção, métodos de Krylov cujos os precondicionadores são eles próprios métodos de Krylov, podendo até ser o mesmo. Esses métodos também são conhecidos como métodos com iterações aninhadas ou com iterações internas-externas . A força dessas alternativas reside no aumento contínuo do subespaço de Krylov. Através da combinação criteriosa do subespaço da iteração interna com o da iteração externa, consegue-se a convergência do método em um número finito de iterações. Essa propriedade não é garantida, por exemplo pelos métodos com recomeço que destroem os subespaços a cada recomeço. Ou seja, restringindo-se os precondicionadores variáveis a métodos de Krylov, garante-se que a iteração global permaneça dentro de um subespaço de Krylov maior e, graças ao teorema 1.2 da pág. 17, temos a convergência. Só que devido ao grau do polinômio mínimo do lado direito b em relação à A, essa convergência pode

ser lenta, sendo necessários precondicionadores, também, para esses métodos.

A relação de Arnoldi (2.1.3), quando do uso de um precondicionador constante pela direita, torna-se

$$AM^{-1}V_n = V_{n+1}\overline{H}_n.$$

Quando o precondicionamento flexível é utilizado, a cada iteração resolvemos o problema $M_i z_i = v_i$, ou seja $z_i = M_i^{-1}v_i$[6]. Seja

$$Z_n = \begin{pmatrix} M_1^{-1}v_1 & \cdots & M_n^{-1}v_n \end{pmatrix}$$

então a relação de Arnoldi escreve-se como

$$AZ_n = V_{n+1}\overline{H}_n, \tag{5.3.5}$$

e se torna necessário estocar ambas as matrizes, Z_n e V_{n+1}, ou seja, aumenta-se o espaço necessário. Nesse caso, a solução aproximada, calculada no k-ésimo passo externo do método, será tal que $x_k - x_0 \in Z_k$, ou, $x_k = x_0 + Z_k u_k$, com $u_k \in \mathbb{R}^k$. Seguindo a sugestão apresentada em [111], vamos alterar a notação para diferenciar as matrizes envolvidas nas iterações internas e externas, assim (5.3.5) passa a

$$AZ_k = W_{k+1}\overline{T}_k, \tag{5.3.6}$$

para nos referirmos as matrizes construídas até k-ésima iteração externa, nesse caso

$$W_k = \begin{pmatrix} w_1 & \cdots & w_k \end{pmatrix}$$

contém a base ortonormal para o espaço externo e $\overline{T}_k$, no caso do uso do método de Arnoldi para ortogonalizar a matriz, contém, em cada coluna i, as coordenadas dos produtos Az_i escritos na base

$$W_{i+1} = \begin{pmatrix} w_1 & \cdots & w_i & w_{i+1} \end{pmatrix}$$

e é uma matrix de Hessenberg superior.

Para evitar confusões, vamos chamar de **ciclo interno** o conjunto de iterações completas do método interno, até convergência ou até a

[6]Entender essa operação como a solução do problema e não, necessariamente, como a inversão da matriz M, pois M pode ser um operador qualquer.

sua parada, e de **ciclo externo** o conjunto de iterações completas do método externo, até convergência ou até a sua parada. Com isso queremos frisar, mais uma vez, a diferença entre uma iteração e um ciclo. Em geral um ciclo será composto de várias iterações, excluído o caso de convergência na primeira iteração dos métodos interno ou externo.

Vamos reescrever o método externo usando a notação proposta em (5.3.6). Dada a aproximação inicial x_0, então temos

$$r_0 = b - Ax_0 \quad w_1 = r_0/\beta \quad \beta = \|r_0\|, \tag{5.3.7}$$

para cada iteração externa k, um vetor z_k, que aproxima a solução de

$$Az = w_k, \tag{5.3.8}$$

é calculado usando um ciclo de um método de Krylov como precondicionador. Em seguida, o vetor Az_k é calculado e ortogonalizado, por um método de Arnoldi, em relação aos vetores anteriores w_i, $i \leq k$ e obtém-se o novo vetor w_{k+1}, com isso o resíduo da k-ésima iteração externa vale

$$r_k = b - Ax_k = r_0 - AZ_k u_k = r_0 - W_{k+1}\overline{T}_k u_k = W_{k+1}(\beta e_1 - \overline{T}_k u_k). \tag{5.3.9}$$

Como estamos usando um método de Arnoldi em cada iteração externa, então, $W_k W_k^T$ e $I - W_k W_k^T$ serão projetores ortogonais em $\mathrm{Im}(W_k)$ e $\mathrm{Im}(W_k)^\perp$, respectivamente. Com isso, ortogonalizar Az_k em relação aos vetores ortonormais w_i, $i \leq k$, é equivalente a deflacionar o vetor resíduo interno $w_k - Az_k$ em relação ao espaço $\mathrm{Im}(W_k)$; vejamos a comprovação dessa afirmação

$$\begin{aligned} t_{k+1,k} w_{k+1} &= (I - W_k W_k^T) A z_k & (5.3.10)\\ &= w_k - (w_k - Az_k) - w_k + W_k W_k^T (w_k - Az_k) \\ &= -(I - W_k W_k^T)(w_k - Az_k). & (5.3.11) \end{aligned}$$

Na k-ésima iteração do método externo um novo vetor z_k, que aproxima (5.3.8), é calculado usando um ciclo interno de Krylov. O subespaço de Krylov construído para estabelecer essa aproximação é

$\mathcal{K}_n(A, w_k)$, com $n = n_k$, e a base desse subespaço pode ser denotada $\{v_1^{(k)}, \ldots, v_n^{(k)}\}$, a qual também atende uma relação do tipo (5.3.5)[7]

$$AV_n^{(k)} = V_{n+1}^{(k)} \overline{H}_n^{(k)}, \tag{5.3.12}$$

e, assim, temos

$$z_k = V_n^{(k)} y_k, \quad \text{para} \quad y_k \in \mathbb{R}^k, \; n = n_k \tag{5.3.13}$$

Outro fato relevante, é que se o método interno for precondicionado pela direita, essa alternativa pode ser vista como uma estratégia de precondicionamento global. Vejamos os detalhes. Suponhamos que se use, no ciclo interno, como precondicionador pela direita fixo uma matriz P, com isso o sistema interno a ser resolvido passa a ser $AP^{-1}\hat{z} = w_k$, com $z = P^{-1}\hat{z}$. O novo espaço de Krylov interno será $\mathcal{K}_n(AP^{-1}, w_k)$. Temos que $\hat{z}_k = V_n^{(k)} y_k$, aonde $V_n^{(k)}$ é uma base para $\mathcal{K}_n(AP^{-1}, w_k)$. Como P é fixo podemos escrever

$$Z_k = P^{-1}\hat{Z}_k = \begin{pmatrix} P^{-1}\hat{z}_1 & \cdots & P^{-1}\hat{z}_k \end{pmatrix}.$$

E a relação de Arnoldi externa (5.3.6) passa a ser

$$AP^{-1}\hat{Z}_k = W_{k+1}\overline{T}_k.$$

Daí podemos concluir que o precondicionamento interno equivale ao precondicionamento externo com a mesma matriz, ainda, pela direita. E assim, usar um precondicionador constante e pela direita no ciclo interno equivale a usá-lo no ciclo externo. Temos então que o método flexível, neste caso, passa a resolver o problema $AP^{-1}\hat{x} = b$, com $\hat{x} = Px$. Podemos considerar, dessa forma, que a matriz A representa a matriz AP^{-1}, não havendo necessidade de sempre explicitar o precondicionador interno.

A seguir, apresentamos o resumo de alguns dos resultados mais relevantes demonstrados em [111]:

[7] Em geral, os métodos de Krylov internos e externos não precisam usar necessariamente o procedimento de Arnoldi, mas nessa seção sempre estaremos considerando que as construções das bases interna e externa se dão através desse procedimento, pois estamos tratando de variantes do GMRES.

1. Relembrando a relação 5.3.6, $AZ_k = W_{k+1}\overline{T}_k$, apesar de Z_k e W_{k+1} não serem subespaços de Krylov, assumindo que eles tenham posto completo, eles pertencem a um espaço de Krylov de dimensão mais elevada, os detalhes e demonstração desse resultado estão no lema 2.2, na pág. 2223.

2. Outro resultado afirma que o subespaço aonde as aproximações são escolhidas continuam crescendo, dadas algumas hipóteses razoáveis, levando à convergência dos métodos com iterações internas-externas em até, no máximo, a ordem da matriz A. Esse resultado é apresentado no teorema 5.2, na página 2228

3. Quanto à estagnação e a ruptura dos métodos, no teorema 6.1 da página 2230, são apresentados resultados sobre as condições e como evitar ruptura e estagnação dos métodos.

5.4 Inexatos

As referências básicas para essa seção são o relatório técnico de Amina Bouras e Valérie Frayssé [19], de 2000, o artigo de Luc Giraud e outros [52], de 2007, os de Valeria Simoncini e Daniel B. Szyld [112] e [114], e o de Jasper van den Eshof e Gerard L. G. Sleijpen [129], de 2004.

Em [19], as autoras propõe a seguinte questão[8]: para métodos de Krylov, qual a melhor estratégia de parada das iterações internas visando assegurar a convergência da iterações externas ao mesmo tempo que o custo computacional global é minimizado? Essa questão foi tratada nos anos 1980 e 1990, no contexto dos métodos de Newton, e a conclusão era de que as iterações internas precisariam de uma maior acurácia quando o processo externo chegasse proximo à convergência. Baseadas em extensa experimentação numérica, as autoras indicam que para métodos de Krylov combinados, com iterações internas-externas, tanto de Arnoldi quanto de Lanczos, a acurácia dos primeiros vetores seria necessária, mas que essa condição poderia ser

[8] Esse trabalho foi publicado, como artigo, em [20].

relaxada com o avanço da convergência do método externo. Esse comportamento, que vai contra a intuição, baseada nas experiências com os métodos de Newton, causou interesse e vários trabalho seguiram a esse primeiro [52], [112], [114], [129], fornecendo a teoria necessária para a compreensão desse fato para diversos métodos de Krylov.

Por exemplo, uma aplicação natural dessa ideia ocorre em eletromagnetismo computacional, aonde o método multipolos rápido fornece aproximações do produto matriz-vetor de acordo com a acurácia definida pelo usuário, quanto menos acurado, mais rápido é o cálculo. O ponto chave é conceber um critério para controlar a acurácia do produto matriz-vetor visando uma convergência satisfatória da iteração. Outro exemplo, vem da área de decomposição de domínio sem recobrimento, aonde o produto matriz-vetor envolvendo a matriz do complemento de Schur (por exemplo, $\Gamma - DA^{-1}G$, originária da matriz $\begin{pmatrix} A & G \\ D & \Gamma \end{pmatrix}$) pode ser aproximada ao invés de calculada exatamente, uma vez que o cálculo exato de A^{-1} pode ser muito caro, tornado-o desaconselhável.

O critério proposto em [19] baseia-se em algumas considerações heurísticas e o procedimento recebeu o nome de estratégia de relaxação porque o tamanho da perturbação cresce inversamente proporcional à norma do resíduo. Denomina-se "GMRES com relaxação", o método GMRES inexato que implementa uma estratégia de relaxação. A estratégia de relaxação proposta em [19] busca assegurar a convergência da iteração do GMRES, x_k, é controlado através de cotas para o erro inverso:

$$\begin{aligned} \eta(x_k) &= \min_{\Delta A,\, \Delta b} \big\{ \tau > 0 : \|\Delta A\| \leq \tau \|A\|, \quad \|\Delta b\| \leq \tau \|b\| \\ &\qquad \text{e} \quad (A + \Delta A)x_k = b + \Delta b \big\} \\ &= \frac{\|Ax - b\|}{\|A\|\|x_k\| + \|b\|} < \varepsilon, \quad \text{para} \quad \varepsilon > 0. \end{aligned} \tag{5.4.14}$$

Baseados em [19], estratégias semelhantes foram aplicadas com sucesso na solução de problemas de difusão heterogênea usando decomposição de domínio [21], como precondicionadores de problemas de

difusão de radiação [135], em problemas de eletromagnetismo [79], em cromodinâmica quântica [40] e em modelos de circulação oceânica de fluxos barotrópicos estáveis [130]. Passos significativos, em direção a uma explicação teórica do comportamento observado acima, foram propostos em [112], [129], e, mais recentemente, em [52]. Nesses trabalhos, são apresentadas justificativas relevantes para o fato de que, garantidas algumas hipóteses razoáveis, o GMRES inexato converge em relação à norma do resíduo.

A convergência de métodos iterativos é usualmente baseada em critérios de erro inverso em relação à norma, ver [11], [43] e [59]. Em [52], são propostos critérios para o controle da acurácia do produtos matriz-vetor e prova-se que eles garantem a convergência do GMRES tanto em relação à $\eta(x_k)$, definido em (5.4.14), quanto a

$$\begin{aligned}\eta_b(x_k) &= \min_{\Delta b} \left\{\tau > 0 : \|\Delta b\| \leq \tau \|b\| \quad \text{e} \quad Ax_k = b + \Delta b\right\} \\ &= \frac{\|Ax - b\|}{\|b\|}. \end{aligned} \tag{5.4.15}$$

Tanto $\eta(x_k)$ quanto $\eta_b(x_k)$ são recomendados em [11] quando se discutem critérios de parada. O critério $\eta_b(x_k) < \varepsilon$ é mais simples do que o critério $\eta(x_k) < \varepsilon$, no entanto os dois tem sua relevância. Sabendo-se que critérios de parada são totalmente dependentes do problema a ser resolvido e da aplicação utilizada, caso as incertezas venham principalmente do lado direito b, então $\eta_b(x_k)$ tem que ser usado, caso venham da matriz e do lado direito então a opção correta é $\eta(x_k)$.

Vamos assumir que seja possível monitorar a acurácia do produto matriz-vetor Av no procedimento de Arnoldi. De um ponto vista matemático, a inacurácia pode ser modelada pela introdução de uma matriz de perturbação E, dependendo possivelmente de v, tal que $(A + E)v$ passe a ser a quantidade realmente calculada. No passo k do algoritmo de Arnoldi perturbado, o vetor $w = (A + E_k)v_k$ é ortogonalizado em relação aos vetores v_j, $j = 1 : k$, com isso temos a seguinte relação:

$$\begin{pmatrix}(A + E_1)v_1 & \cdots & (A + E_k)v_k\end{pmatrix} = \begin{pmatrix}v_1 & \cdots & v_k & v_{k+1}\end{pmatrix} \overline{H}_k, \tag{5.4.16}$$

onde $V_{k+1} := \begin{pmatrix}v_1 & \cdots & v_k & v_{k+1}\end{pmatrix}$ é uma matriz com colunas orto-

normais e $\overline{H}_k$ é uma matriz Hessenberg superior. Vamos assumir que (assim como em [52], [112] e [129]) o produto matriz-vetor que ocorre no cálculo do resíduo inicial é exato, ou seja $r_0 = b - Ax_0$ e $\beta = \|b - Ax_0\|$. Definamos a k-ésima iteração do método inexato como sendo $x_k = x_0 + \delta x_k$, onde $\delta x_k = V_k y_k$ e y_k é a solução do problema de quadrados mínimos linear $\min_y \|\beta e_1 - \overline{H}_k y\|$. Introduzindo a matriz de perturbação $G_k = \begin{pmatrix} E_1 v_1 & \cdots & E_k v_k \end{pmatrix}$, o problema inexato de Arnoldi pode ser escrito como um problema exato de Arnoldi para um problema aproximado:

$$\widetilde{A}_k V_k = (A + G_k V_k^T) V_k = V_{k+1} \overline{H}_k, \quad \text{onde} \quad \widetilde{A}_k = A + G_k V_k^T.$$

A última igualdade mostra que as quantidades δx_i, $\overline{H}_i$ e v_i para $i \leq k$ geradas pelo GMRES inexato de $A\delta x = r_0$ até ao passo k são as mesmas que as geradas pelos primeiros k passos do GMRES exato aplicado ao sistema linear $\widetilde{A}_k \delta x = r_0$. Usando alguns resultados clássicos sobre o GMRES exato [101], o resultado anterior implica, por indução, que a norma do resíduo $r_0 - \widetilde{A}_k \delta x_i$ é monotonamente decrescente com o crescimento de i, com $i \leq k$. E mais, $\widetilde{A}_k \delta x_i = \widetilde{A}_i \delta x_i$ pois

$$\widetilde{A}_k \delta x_i = (A + G_k V_k^T) V_i y_i = (A + G_i V_i^T) V_i y_i = \widetilde{A}_i \delta x_i.$$

Logo, a norma do resíduo calculado, $\widetilde{r}_k = r_0 - \widetilde{A}_k \delta x_k$, diminui com k. Vamos chamar de r_k o resíduo $r_0 - A\delta x_k$ e seja $\widetilde{r}_0 = r_0$. Então, a iteração inexata de Arnoldi também pode ser escrita

$$\widetilde{A}_k V_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^T.$$

Baseando-nos na terminologia usada para o GMRES exato, diremos que uma ruptura ocorre no passo n se $h_{n+1,n} = 0$. Por causa da ortogonalidade de V_k, assim como para o GMRES exato, a ruptura irá ocorrer para $n \leq m$, onde m é a ordem de A. A cada passo, a **separação residual** é definida como $r_k - \widetilde{r}_k = (A - \widetilde{A}_k)\delta x_k = G_k y_k = \sum_{i=1}^{k} y_{k,i} E_i v_i$, onde $y_k = \begin{pmatrix} y_{k,1} & \cdots & y_{k,k} \end{pmatrix}^T \in \mathbb{R}^k$. Em [52] são demonstradas desigualdades para se estimar as separações residuais de cada passo.

A partir dessa nomenclatura é possível discutir critérios para a convergência dos GMRES inexato. Em [52], são apresentados quatro resultados:

1. Dadas algumas condições de controle das perturbações permitidas, a ruptura do GMRES inexato é sempre benéfica, teorema 1, pág. 714,

2. Apresenta condições para $\eta_b(x_k)$ ser menor do que qualquer tolerância prescrita no teorema 2, pág. 715,

3. Apresenta condições para $\eta(x_k)$ ser menor do que qualquer tolerância prescrita no teorema 3, pág. 716,

4. Apresenta condições para convergência, em relação à $\eta(x_k)$, no caso em que mesmo o primeiro produto matriz-vetor é inexato, pág. 717, o que ocorre em situações em que a matriz é necessariamente aproximada, como para matrizes do complemento de Schur de grande ordem.

Todos os resultados e análises feitos até agora dizem respeito ao GMRES inexato sem a alternativa de recomeço, o que dificulta o uso prático do método, uma vez que o recomeço é essencial. No entanto, em [52], os autores apresentam um resultado de estimativa do erro direto em relação ao erro inverso e ao condicionamento da matriz A para o GMRES inexato com recomeço.

Cabe enfatizar que os resultados aqui comentados, assim como os dos outros artigos referenciados, baseiam-se no conhecimento do menor valor singular da matriz A, que pode ser um cálculo difícil e caro. Sendo assim é necessário o estabelecimento de resultados de convergência utilizando parâmetros mais simples e econômicos, sendo essa uma questão em aberto.

Vale também o alerta apresentado em [112, pág. 455], onde os autores mostram que mesmo ocorrendo convergência do GMRES inexato, a taxa de convergência pode piorar em relação à do GMRES exato e apresentam exemplo, na página 465 do referido artigo, para caracterizar uma situação com a convergência degradada.

Exercícios

1. No passo 3: do algoritmo 5, deve-se calcular k vetores harmônicos de Ritz. Esse vetores devem ser calculados em relação a que espaços? Qual fórmula deve ser usada?

2. Estude e faça os detalhes do teorema 3.2 em [85, pág. 25].

3. Estude e faça os detalhes do teorema 3.3 em [85, pág. 26].

4. Faça a implementação em Matlab, ou equivalente, do GMRES-DR.

5. Para o GMRES-DR, faça comparações entre várias escolhas dos k valores harmônicos de Ritz: menores, maiores, randômicos.

6. Na página 86 podemos ler a afirmação que quando uma matriz "tem autovalores equidistantes da origem, ou seja pertencem a um círculo cuja o centro é a origem, o GMRES ficará estagnado por um número de passos igual ao número de autovalores que estão nesse círculo, independente do tamanho dos autovalores." Usando um código de GMRES qualquer (de preferência o que o leitor já implantou) teste essa afirmação. Faça para autovalores com diferentes módulos.

7. Na página 86 encontra-se a afirmação de que: para uma matriz "se os autovalores estão agregados em torno da origem, pode se tornar impossível até mesmo calcular esse autovalores, e suas aproximações, com acurácia suficiente." Usando um código de GMRES qualquer (de preferência o que o leitor já implantou) teste essa afirmação, para matrizes com autovalores pequenos, ou seja, bem próximos do zero da máquina.

8. Implemente a alternativa de GMRES inexato apresentada em [18].

9. Teste o exemplo sugerido em [112, pág. 465] para o GMRES inexato.

Capítulo 6

Estudo de Caso: GMRES Flexível com Recomeço Deflacionado

A solução de sistemas lineares de grande porte é um dos núcleos essenciais de simulações industriais e científicas de larga escala e os MPSK precondicionados estão entre os solvers mais populares. Como falamos anteriormente, para matrizes não simétricas o GMRES [102] é frequentemente escolhido devido a sua robustez. Como apontamos na seção 2.3, há trabalhos sobre o tema em [91] e [96] onde são caracterizadas a estabilidade em relação ao erro inverso das implementações do GMRES, usando as reflexões de Householder e o método modificado de Gram-Schmidt no processo de Arnoldi. Uma outra razão da popularidade do método é que a norma euclidiana do resíduo não cresce (em geral decresce) durante o avanço das iterações. No entanto, para fazer do GMRES uma alternativa factível duas propriedades devem ser combinadas: o uso parcimonioso da memória disponível e o número de operações realizadas deve ser pequeno. Para tanto, algum processo de recomeço do GMRES é necessário, uma vez que, com o avançar das iterações, as duas propriedades são perdidas. Na abordagem clássica de recomeço, apenas um vetor é guardado, de forma a garantir que o resíduo continue sua diminuição (ou, pelo menos, seu não aumento).

Como vimos no capítulo 5, tem sido observado que a reutilização de alguns vetores do espaço de Krylov calculado anteriormente, e não apenas da melhor solução aproximada, para a construção dos espaços necessários à proxima iteração pode ter um impacto positivo na convergência do método. Em várias abordagens, alguma estimativa dos espaços invariantes é recuperada no subespaço de Krylov em uso e reutilizada para o novo recomeço:

1. aumentando o subespaço [27], [83], [100],
2. fazendo uma a ortogonalidade em relação a uma parte relevante do subespaço anterior [92].

Como discutimos na seção 5.1, foi apresentada uma versão do GMRES com deflação GMRES-DR em [85]. Essa alternativa reduz-se ao próprio GMRES quando nenhuma deflação é utilizada, mas pode proporcionar uma convergência bem mais rápida do que o GMRES para exemplos acadêmicos, caso haja uma escolha criteriosa dos espaços de deflação, ver [85].

Uma característica comum a todos os métodos citados anteriormente é que eles se baseiam em um precondicionador fixo M, que será usado reiteradamente durante todo o processo. Há, no entanto, situações aonde essa condição não pode ser atendida. Um exemplo ocorre no uso de técnicas de decomposição de domínio, quando solvers aproximados são necessários para a solução dos problemas interiores, ver [117, sec. 4.4], [125, sec. 4.3]. Esse procedimento se faz necessário quando os problemas locais tornam-se muito grandes para serem resolvidos por métodos diretos e algum solver iterativo é chamado. Se o precondicionador baseado em decomposição de domínio faz uso de solvers aproximativos, métodos como o GMRES flexível são adequados, ver seção 5.3.

Esse capítulo discutirá um novo método proposto em [53] e está baseado, principalmente, nesse trabalho. Esse método combina as iterações flexíveis com uma estratégia de recomeço que recupera informação sobre autoespaços aproximados que estão disponíveis ao fim do ciclo anterior. O sistema a ser resolvido, $Ax = b$, está definido sobre o corpo dos números complexos. $A \in \mathbb{C}^{m\times m}$ é regular, b e x estão em

$\mathbb{C}^m$. Em resumo, o método começa com um vetor inicial $x_0 \in \mathbb{C}^m$ e busca soluções aproximadas x_k tal que $x_k - x_0 \in \mathcal{K}_k$. Seja $V_k \in \mathbb{C}^{m\times k}$ uma matriz cujas colunas formam uma base ortonormal para $\mathcal{K}_k$. Essa matriz será construída por um método de Arnoldi e gozará da seguinte relação

$$AV_k = V_{k+1}\overline{H}_k,$$

com a condição que $r_0 := b - Ax_0 \in \mathcal{K}_k$. Como antes, $\overline{H}_k \in \mathbb{C}^{(k+1)\times k}$. Esse novo método, buscará a cada iteração minimizar a norma euclidiana do resíduo, ou seja $r_k \perp \mathcal{L}_k$, como faz o GMRES.

6.1 Apresentação do Método

Métodos flexíveis implementam um esquema tal que, após um número fixo de iterações (denotado n nesse capítulo), o subespaço de Krylov é truncado e o método é recomeçado para garantir o controle sobre o uso da memória e diminuir o custo do processo de ortogonalização. No FGMRES, o método é suspenso e reiniciado, tomando-se como novo valor inicial o vetor que permitiu a menor norma euclidiana do resíduo, semelhante ao GMRES com recomeço, a diferença aqui é que os precondicionadores mudam de uma iteração a outra. No caso do GMRES-DR, um esquema mais sofisticado é utilizado: um subespaço especial de dimensão $k < n$ é guardado de uma iteração a outra, além do melhor valor inicial também. Vários exemplos de sucesso desse segundo método, para problemas acadêmicos, foram apresentados em [83]. Mostraremos como estender GMRES-DR de forma a permitir que precondicionadores variáveis sejam incorporados ao método. Vamos denotar por M_i a operação que representa o precondicionamento no passo i do método. No algoritmo 6, na página 102, o algoritmo do método FGMRES-DR é apresentado. Assim como o algoritmo do método FGMRES, começando de uma estimativa inicial x_0, ele gera, a cada recomeço, as matrizes $Z_n \in \mathbb{C}^{m\times n}$, $V_{n+1} \in \mathbb{C}^{m\times(n+1)}$ e $\overline{H}_n \in \mathbb{C}^{(n+1)\times n}$ tais que $AZ_n = V_{n+1}\overline{H}_n$. Uma solução aproximada x_n é encontrada através da minimização da norma euclidiana do do resíduo $\|b - A(x_0 + V_n y)\|_2$ tal que $x_n - x_0 \in \mathrm{Im}(V_n)$, e o vetor resíduo correspondente $r_n = b - Ax_n \in \mathbb{C}^m$, com $r_n \in \mathrm{Im}(V_{n+1})$ e

Algoritmo 6 FGMRES com recomeço deflacionado

1: *Inicialização:* Escolha $n > 0,\ k > 0,\ tol > 0,\ x_0 \in \mathbb{C}^m$. Sejam $r_0 = b - Ax_0$, $\beta = \|r_0\|_2$, $c = \beta e_1$, $v_1 = r_0/\beta$.

2: *Primeiro recomeço:* Aplique o processo de Arnoldi com precondicionador flexível e construa V_{n+1}, Z_n e $\overline{H}_n$, tal que:
$$AZ_n = V_{n+1}\overline{H}_n.$$

3: *Solução de norma mínima:* Calcule a aproximação x_n tal que $x_n - x_0 \in \mathrm{Im}(Z_n)$, ou seja $x_n - x_0 = Z_n y_n$, onde $y_n = \arg\ \min_{y\in\mathbb{C}^k} \|c - \overline{H}_n y\|_2$. Coloque $x_0 = x_n$ e $r_0 = b - Ax_0$.

4: *Teste de convergência:* Se $\|c - \overline{H}_n y_n\|_2 \leq tol$. Fim.

5: *Procedimento de recomeço:* Faça fatoração QR de $\{u_1, \ldots, u_k\}$ e coloque o fator Q em V_k^{nova}, onde $u_i \in \mathbb{C}^n$, $i = 1 : k$, são os vetores harmônicos de Ritz de $AZ_m V_m^H$ em relação à $\mathrm{Im}(V_m)$ (veja algoritmo 7, pág.103). Use Gram-Schmidt modificado para obter v_{k+1}^{nova}, tal que $r_0 \in \mathrm{Im}(V_{k+1}^{\text{nova}})$, onde $V_{k+1}^{\text{nova}} = \begin{pmatrix} V_k^{\text{nova}} & v_{k+1}^{\text{nova}} \end{pmatrix}$ é ortonormal. Calcule Z_k^{nova} e $\overline{H}_k^{\text{nova}}$ de modo que
$$AZ_k^{\text{nova}} = V_{k+1}^{\text{nova}}\overline{H}_k^{\text{nova}}. \tag{6.1.1}$$

6: *Laço interno:* Aplique $(m - k)$ passos adicionais do método de Arnoldi usando um precondicionamento flexível em (6.1.1), para chegar a:
$$AZ_n^{\text{nova}} = V_{n+1}^{\text{nova}}\overline{H}_n^{\text{nova}}. \tag{6.1.2}$$

7: *Procedimentos para recomeço:* $c = \left(V_{n+1}^{\text{nova}}\right)^H r_0$, $Z_n = Z_n^{\text{nova}}$, $V_{n+1} = V_{n+1}^{\text{nova}}$, $\overline{H}_n = \overline{H}_n^{\text{nova}}$. Recomece em 3.

$r_n \perp \mathrm{Im}(AV_n)$. O único passo que se mantém sem especificação é o procedimento de recomeço, onde Z_k^{nova}, V_{k+1}^{nova} e $\overline{H}_k^{\text{nova}}$ são calculados de tal forma que a equação (6.1.1) seja válida. O teorema 6.1 e o algoritmo 7 mostram como essas alternativas podem ser implantadas com eficiência. O fundamento dessa abordagem é o uso de certos vetores harmônicos de Ritz. Relembrando. Seja um espaço $\mathcal{C} \subset \mathbb{C}^m$ e seja C uma matriz cujas colunas contêm uma base para esse espaço. Relembrando o conceito da seção 4.1. Seja $B \in \mathbb{C}^{m\times m}$ uma matriz. O par $(y, \theta) \in \mathcal{C} \times \mathbb{C}$ é um par harmônico de Ritz de B em relação a BC se e somente se (ver teorema 4.4, na pág. 64):

$$C^H B^H (By - \theta y) = 0, \tag{6.1.3}$$

onde y é um vetor harmônico de Ritz associado ao valor harmônico de Ritz θ.

Antes de demonstrarmos algumas das propriedades do método, faremos dois comentários.

Algoritmo 7 Cálculos de Z_k^{nova}, V_{k+1}^{nova} e $\overline{H}_k^{\text{nova}}$

1: *Entradas:* A, Z_n, V_{n+1} e $\overline{H}_n$, tais que $AZ_n = V_{n+1}\overline{H}_n$.

2: *Calculo de k vetores harmônicos de Ritz:* Calcule k autovetores independentes g_i da matriz $H_n + h_{(n+1),n}^2 H_n^{-H} e_n e_n^H$ com $e_n^H = (0_{n-1}, 1)$, onde 0_{n-1} é o vetor-linha nulo $1 \times (n-1)$ e H_n são a primeiras n linhas de $\overline{H}_n$. Coloque $G_k = \begin{pmatrix} g_1 & \cdots & g_k \end{pmatrix} \in \mathbb{C}^{n \times k}$.

3: *Aumento de G_k:* Coloque

$$\begin{pmatrix} \begin{pmatrix} G_k \\ 0_k^T \end{pmatrix} & c - \overline{H}_n y_m \end{pmatrix}$$

onde 0_k^T e o vetor-linha nulo $1 \times k$.

4: *Ortonormalização das colunas de G_{k+1}:* Faça a fatoração QR de $G_{k+1} = P_{k+1}\Gamma_{k+1}$ e armazene P_{k+1}, $P_k = P_{k+1}(1:k, 1:k)$.

5: Coloque $V_{k+1}^{\text{nova}} = V_{n+1}P_{k+1}$, $Z_k^{\text{nova}} = Z_n P_k$ e $\overline{H}_k^{\text{nova}} = P_{k+1}^H \overline{H}_n P_k$.

Observação 6.1 (Ruptura do algoritmo)**.** *O passos 2: e 6: do algoritmo 6 implementam o algoritmo FGMRES sem recomeço, mas com um número máximo de iterações, o qual pode sofrer uma ruptura não benéfica antes que a solução do sistema linear tenha sido encontrada, veja [99]. Uma ruptura ocorre no passo j quando $AM_j v_j$ pertence a* $\text{Im}(V_j)$*; essa situação corresponde a uma propriedade particular de M_j com relação a A e V_j, que, de nosso ponto de vista, não deverá ocorrer quando M_j significa a utilização de um método iterativo precondicionado. Devemos lembrar, usando a teoria desenvolvida no capítulo 2, que para o GMRES essa situação não ocorrerá, uma vez que para esse método apenas ocorrem rupturas benéficas, caso o precondicionador M_j seja constante. Um outro ponto a ser observado, é que o FGMRES-DR se baseia no cálculo de k autopares distintos de uma matriz no passo 2: do algoritmo 7. Isso pode não ser possível caso a matriz não seja diagonalizável, mas como vimos no teorema 7, na pág. 103, esse fato é raro em precisão finita, pois uma leve perturbação transforma uma matriz qualquer em uma matriz diagonalizável. Podemos estimar, então, que será pouco provável que um método de cálculos de autopares como a fatoração QR não consiga encontrar k autovetores linearmente independentes da matriz dada. Ou seja, o método proposto FGMRES-DR, assim como seu precursor FGMRES, pode sofrer uma ruptura não benéfica, mas podemos considerar essa possibilidade remota em aplicações práticas.*

Observação 6.2 (Uso de valores harmônicos de Ritz)**.** *No passo 5: do algoritmo 6 calculamos pares harmônicos de Ritz associados à matriz $AZ_nV_n^H$. A versão precondicionada do método GMRES-DR calcula, no seu procedimento de recomeço, os vetores harmônicos de Ritz de AM em relação à V_n. Da equação* (6.1.3) *podemos concluir que esses vetores são os vetores harmônicos de Ritz de $AMV_nV_n^H = AZ_nV_n^H$ em relação à V_n que é o resultado que usamos no FGMRES-DR. Podemos então considerar que o procedimento de recomeço do algoritmo 6 pode ser visto como uma generalização da operação correspondente no método FGMRES-DR, quando do uso de precondicionadores flexíveis.*

Passamos à demonstração do teorema que mostra que uma relação do tipo Arnoldi é garantida pelo método.

Teorema 6.1. *Seja P_k a matriz definida no algoritmo 7. A cada recomeço do FGMRES-DR a seguinte relação de Arnoldi é valida*

$$AZ_k^{nova} = V_{k+1}^{nova}\overline{H}_k^{nova}, \tag{6.1.4}$$

com

$$Z_k^{nova} = Z_nP_k, \tag{6.1.5}$$

$$V_{k+1}^{nova} = V_{n+1}P_{k+1} \tag{6.1.6}$$

e

$$\overline{H}_k^{nova} = P_{k+1}^H\overline{H}_nP_k \tag{6.1.7}$$

Demonstração: Para a matriz de Hessenberg $\overline{H}_n$, vamos calcular seus pares harmônicos de Ritz. Usando o teorema 4.6 da pág. 68, que relaciona os pares harmônicos com o resíduo de uma iteração, podemos afirmar que existem $\alpha_i \in \mathbb{C}$,

$$\overline{H}_ng_i - \lambda_i\begin{pmatrix} g_i \\ 0 \end{pmatrix} = \alpha_i(c - \overline{H}_ny_n) = \alpha_i\rho_n, \quad i = 1:k,$$

com $\rho_n := c - \overline{H}_ny_n$. Multiplicando à esquerda por V_{n+1} temos

$$V_{n+1}\overline{H}_ng_i - \lambda_iV_{n+1}\begin{pmatrix} g_i \\ 0 \end{pmatrix} = V_{n+1}\alpha_i\rho_n, \quad i = 1:k,$$

e por sua vez, usando a relação de Arnoldi $AZ_n = V_{n+1}\overline{H}_n$,

$$AZ_n g_i = V_{n+1}\left(\lambda_i \begin{pmatrix} g_i \\ 0 \end{pmatrix} + \alpha_i \rho_n\right), \quad i = 1:k. \tag{6.1.8}$$

Colocando $G_k = \begin{pmatrix} g_1 & \cdots & g_k \end{pmatrix} \in \mathbb{C}^{n\times k}$ e $\alpha = \begin{pmatrix} \alpha_1 & \cdots & \alpha_k \end{pmatrix} \in \mathbb{C}^{1\times k}$, a equação (6.1.8) passa a ser

$$AZ_n G_k = V_{n+1}\left(\begin{pmatrix} G_k \\ 0_k^T \end{pmatrix} \quad \rho_n\right)\begin{pmatrix} \operatorname{diag}(\lambda_1,\ldots,\lambda_k) \\ \alpha \end{pmatrix}. \tag{6.1.9}$$

Façamos a fatoração QR de $G_k = P_k\Gamma_k$ e ortogonalizemos ρ_n em relação às colunas de $\begin{pmatrix} G_k \\ 0_k^T \end{pmatrix}$ para obtermos p_{k+1}. Temos que

$$ap_{k+1} = \rho_n - \begin{pmatrix} P_k \\ 0_k^T \end{pmatrix}^H u,$$

onde $a = \left\|\rho_n - \begin{pmatrix} P_k \\ 0_k^T \end{pmatrix} u\right\|$ e $u_i = \begin{pmatrix} p_i \\ 0 \end{pmatrix}\rho_n$. A matriz ortogonal P_{k+1} pode ser escrita como

$$P_{k+1} = \left(\begin{pmatrix} P_k \\ 0_k^T \end{pmatrix} \quad p_{k+1}\right).$$

Em relação ao fator Γ_{k+1}, temos que

$$\begin{aligned}\left(\begin{pmatrix} G_k \\ 0_k^T \end{pmatrix} \quad \rho_n\right) &= \left(\begin{pmatrix} G_k \\ 0_k^T \end{pmatrix} \quad ap_{k+1} + \begin{pmatrix} P_k \\ 0_k^T \end{pmatrix} u\right) \\ &= \left(\begin{pmatrix} P_k \\ 0_k^T \end{pmatrix} \quad p_{k+1}\right)\begin{pmatrix} \Gamma_k & u \\ 0_k^T & a \end{pmatrix},\end{aligned}$$

logo

$$\Gamma_{k+1} = \begin{pmatrix} \Gamma_k & u \\ 0_k^T & a \end{pmatrix}.$$

Como consequência, a equação (6.1.9) torna-se

$$AZ_n G_k = V_{n+1}P_{k+1}\Gamma_{k+1}\begin{pmatrix} \operatorname{diag}(\lambda_1,\ldots,\lambda_k) \\ \alpha \end{pmatrix}\Gamma_k^{-1}.$$

Devemos notar que Γ_k é regular se $\dim(\mathrm{Im}(G_k)) = k$, mas este fato está sendo assumido, ver observação 6.1 sobre as possíveis causas de ruptura do algoritmo.

Denotemos $Z_k^{\text{nova}} = Z_n P_k$, $V_{k+1}^{\text{nova}} = V_{n+1} P_{k+1}$ e

$$\overline{H}_k^{\text{nova}} = \Gamma_{k+1} \begin{pmatrix} \mathrm{diag}(\lambda_1, \ldots, \lambda_k) \\ \alpha \end{pmatrix} \Gamma_k^{-1},$$

dessa forma podemos escreve a relação de Arnoldi

$$AZ_k^{\text{nova}} = V_{k+1}^{\text{nova}} \overline{H}_k^{\text{nova}}.$$

Agora vamos escrever $\overline{H}_k^{\text{nova}}$ como um produto de matrizes com ordens pequenas (basicamente k e n). A relação de paralelismo

$$\overline{H}_n g_i - \lambda_i \begin{pmatrix} g_i \\ 0 \end{pmatrix} = \alpha_i \rho_n, \quad i = 1 : k,$$

pode ser colocada em forma matricial como

$$\overline{H}_n G_k - \begin{pmatrix} G_k \\ 0_k^T \end{pmatrix} \mathrm{diag}(\lambda_1, \ldots, \lambda_k) = \rho_n \alpha.$$

Como $G_k = P_k \Gamma_k$, podemos ter

$$\overline{H}_n P_k \Gamma_k = P_k \Gamma_k \begin{pmatrix} \mathrm{diag}(\lambda_1, \ldots, \lambda_k) \\ \alpha \end{pmatrix},$$

$$P_k^H \overline{H}_n P_k = \Gamma_k \begin{pmatrix} \mathrm{diag}(\lambda_1, \ldots, \lambda_k) \\ \alpha \end{pmatrix} \Gamma_k^{-1}.$$

E assim $\overline{H}_k^{\text{nova}} = P_k^H \overline{H}_n P_k$ como está colocado no algoritmo 7. ■

Vamos agora discutir sobre c, o lado direito do problema de quadrados mínimos que deve ser resolvido em cada iteração:

$$y_n = \arg \min_{y \in \mathbb{C}^k} \|c - \overline{H}_n y\|_2.$$

No método GMRES-DR, o lado direito do problema de quadrados mínimos é calculado através de $c = V_{n+1}^H r_0$. Esse cálculo requer a

realização de $(n+1)$ produtos internos, com vetores de tamanho m. A primeira observação é que $r_0 \in \mathrm{Im}(V_{k+1})$, que por sua vez é ortogonal a $V_k(:, k+2:n)$, logo, podemos garantir que haverá $(n-k)$ valores nulos na multiplicação para o cálculo de c, logo não precisamos fazê-las, o que reduz o cálculo a $c = V_{k+1}^H r_0$. Explorando um pouco mais essa observação em relação a r_0, o cálculo de c pode ser ainda simplificado, como veremos no próximo teorema.

Teorema 6.2. *A cada recomeço de ciclo, o novo resíduo $r_0 \in V_{k+1}^{nova}$ tem suas coordenadas dadas pela última coluna do fator R da decomposição QR da matriz*

$$\begin{pmatrix} \begin{pmatrix} G_k \\ 0_k^T \end{pmatrix} & c - \overline{H}_n y_n \end{pmatrix}.$$

Demonstração: Seja $P_{k+1}\Gamma_{k+1}$ a fatoração QR de

$$\begin{pmatrix} \begin{pmatrix} G_k \\ 0_k^T \end{pmatrix} & c - \overline{H}_n y_n \end{pmatrix}$$

com $\begin{pmatrix} u \\ a \end{pmatrix}$ sendo a última coluna de Γ_{k+1}. Temos que $c - \overline{H}_n y_n = P_{k+1}\begin{pmatrix} u \\ a \end{pmatrix}$, logo

$$r_0 = V_{n+1}(c - \overline{H}_n y_n) = V_{n+1} P_{k+1} \begin{pmatrix} u \\ a \end{pmatrix} = V_{k+1}^{\mathrm{nova}} \begin{pmatrix} u \\ a \end{pmatrix}.$$

■

6.2 Implementação Computacional

Os resultados apresentados nos teoremas 6.1 e 6.2 nos permitem a formatação simples e efetiva para o método FGMRES-DR, que apresentamos no algoritmo 8. Apesar do método ter sido descrito sobre o corpo dos complexos, ele pode ser usado para reais, apenas separando as partes real e imaginária dos vetores harmônicos de Ritz no

passo 6: do algoritmo 8, como já havia sido sugerido em [85] para o GMRES-DR.

O passo 10: do algoritmo 8 oferece duas possibilidades de cálculo de c, que nada mais é do que o resíduo descrito na base ortonormal formada pelas colunas de V_{n+1}. Como foi sugerido no teorema 6.2, o uso da expressão envolvendo γ_{k+1} é atraente pois evita o cálculo do produto matriz vetor $V_{k+1}^{\text{nova}} r_0$. Ainda cabe um estudo para entender o comportamento dessas duas alternativas na presença de erros de arredondamento.

Quanto ao consumo de memória, como o objetivo desses métodos é a solução de problemas de grande porte, onde m é muito maior do que k e n, para analisar o consumo de memória podemos desconsiderar todos os vetores e matrizes que envolvem apenas as dimensões n e k, mas somos obrigados a tomar cuidado com qualquer operação que envolva a dimensão m. Usando essa convenção, FGMRES é duas vezes mais caro em memória do que GMRES, o mesmo valendo para FGMRES-DR. Na alternativa flexível as bases para os espaços têm que ser armazenadas, tanto V quanto Z (a parte precondicionada). Podemos diminuir o consumo de memória sobrescrevendo Z^{nova} e V^{nova} em Z e V. Isso pode ser conseguido no passo 8: do algoritmo 8, através de multiplicações usando o mesmo espaço de memória. Podemos realizar, também, uma fatoração LU de P_{k+1}, com pivoteamento total para garantir estabilidade, e fazer multiplicações com os fatores triangulares de P_{k+1}. Esse procedimento pode salvar bastante memória e pode ser monitorado graças ao quociente

$$\frac{\|P_k - LU\|}{\|P_k\|}. \tag{6.2.10}$$

Exercícios

1. Construa exemplos de situações de ruptura para os métodos FGMRES e por conseguinte do FGMRES-DR.

2. Faça a implantação em Matlab, ou equivalente, do algoritmo 8.

3. A partir da implantação do item 2 faça as implantações do GMRES-DR e do FGMRES.

4. Implemente a ideia de fazer as multiplicações do passo 8: do algoritmo 8, através de multiplicações usando o mesmo espaço de memória, através da fatoração LU de P_{k+1}, com pivoteamento total. Qual o ganho em termos de armazenamento? Caso seja usado pivoteamento parcial, qual será a diferença? Faça medidas da perda de informação.

5. Na equação (6.2.10) não foi especificada nenhuma norma. Como continuação do exercício anterior, meça a perda de informação com algumas normas diferentes e tente estimar qual seria a melhor.

Algoritmo 8 Implementação do FGMRES-DR: **FGMRES-DR(m, k)**

1: *Inicialização:* Escolha de:

m : a maior dimensão do subespaço para busca de solução,

k : o número de autovetores aproximados,

tol : o limite para a convergência,

x_0 : a aproximação inicial.

Defina $r_0 = b - Ax_0$, $\beta = \|r_0\|_2$, $c = \beta e_1$, $v_1 = r_0/\beta$.

2: *Primeira iteração:* Aplique o FGMRES e gere V_{n+1}, Z_n e $\overline{H}_n$.

3: *Solução com norma mínima:* Resolva $y_n = \arg\min_{y \in \mathbb{C}^k} \|c - \overline{H}_n y\|_2$. Calcule x_n tal que $x_n - x_0 = Z_n y_n$. Calcule o resíduo $r_n = V_{n+1}^H \big(c - \overline{H}_n y_n\big)$. Atribua $x_0 = x_n$ e $r_0 = r_n$.

4: *Teste de convergência:* Se $\|r_0\|_2/\|b\|_2 = \|c - \overline{H}_n y_n\|_2/\|b\|_2 \leq tol$. Fim.

5: *Cálculo dos pares harmônicos de Ritz:* Calcule k autovetores independentes g_i da matriz $H_n + h_{(n+1),n}^2 H_n^{-H} e_n e_n^H$. Armazene g_i, $i = 1 : k$ em G_k.

6: *Agregação dos quase-resíduos $\beta e_1 - \overline{H}_n y_n$ a G_k:*

- *A* **complexa**: agregue um vetor-linha de zeros a G_k e coloque $c - \overline{H}_n y_n$ na ultima coluna.
- *A* **real**: se algum g_i é complexo, separe as partes real e imaginária, caso haja mais do que k vetores, dispense um vetor real. Agregue um vetor-linha de zeros a G_k e coloque $c - \overline{H}_n y_n$ na ultima coluna.

Nos dois casos, a nova matriz G_{k+1} tem dimensões $(n+1) \times (k+1)$.

7: *Ortonormalização das colunas de G_{k+1}:* Faça a fatoração QR de $G_{k+1} = P_{k+1}\Gamma_{k+1}$, armazene P_{k+1} e a última coluna de Γ_{k+1}, γ_{k+1}.

8: *Novas matrizes:* $V_{k+1}^{\text{nova}} = V_{n+1} P_{k+1}$, $Z_k^{\text{nova}} = Z_n P_k$ e $\overline{H}_k^{\text{nova}} = P_{k+1}^H \overline{H}_n P_k$.

9: *Laço interno:* Aplique $(n-k)$ passos adicionais do método de Arnoldi flexível em $AZ_k^{\text{nova}} = V_{k+1}^{\text{nova}} \overline{H}_k^{\text{nova}}$ tal que ao fim se tenha $AZ_n^{\text{nova}} = V_{n+1}^{\text{nova}} \overline{H}_n^{\text{nova}}$.

10: *Recomeço:* Coloque $c = \big(V_{k+1}^{\text{nova}}\big)^H r_0$ e complete com zeros ou $c = \begin{pmatrix} \gamma_{k+1} \\ 0_{n-k} \end{pmatrix}$, $Z_n = Z_n^{\text{nova}}$, $V_{n+1} = V_{n+1}^{\text{nova}}$, $\overline{H}_n = \overline{H}_n^{\text{nova}}$. Recomece em 3. Nesse expressão 0_{n-k} é um vetor nulo de $n-k$ posições.

Apêndice A

Revisão de Álgebra Linear

Os prerrequisitos desse livro são, como informado anteriormente, um curso básico de álgebra linear e a capacidade/disposição de realizar demonstrações matemáticas simples, dedutivas ou indutivas. No entanto, objetivando criar linguagem e base comuns, vamos apresentar alguns dos resultados fundamentais que usaremos no transcorrer dessa obra. Esse apêndice não se propõe completo, logo, algumas definições e demonstrações são assumidas conhecidas, muitas deixadas como exercícios ao final do apêndice, e outras tantas, simplesmente, omitidas. Como obras de referência, que complementam e ultrapassam, de muito, esse resumo, podemos citar [67], [76], [81] e [122].

A.1 Operações

Dado o caráter peculiar e a onipresença da operação de multiplicação de matrizes nesse trabalho, sendo a multiplicação de matriz por vetor um caso particular, vamos tratá-la observando algumas interpretações úteis para as demonstrações. Nessa seção, consideraremos $A \in \mathbb{C}^{m\times p}$, $B \in \mathbb{C}^{p\times n}$ e $C \in \mathbb{C}^{m\times n}$.

A.1.1 Multiplicação: produtos linha por coluna

É o algoritmo básico de multiplicação ensinado nos cursos de graduação:

$$C(i,j) = \sum_{k=1}^{p} A(i,k) \cdot B(k,j) = A(i,:) \cdot B(:,j).$$

Cada elemento de C é produto de uma linha de A por uma coluna de B. Para matrizes reais, esses produtos podem ser considerados como produtos internos. Uma outra leitura relevante é que cada coluna de C se origina na coluna equivalente de B, $C(:,j) = A \cdot B(:,j)$ e que cada uma de suas linhas à equivalente em A, $C(i,:) = A(i,:) \cdot B$.

A.1.2 Multiplicação: produtos externos

Nesse caso a operação completa é representada por apenas uma fórmula

$$C = \sum_{k=1}^{p} A(:,k) \cdot B(k,:).$$

Cada parcela desse somatório é, ela própria, uma matriz $m \times n$; são todas matrizes de posto 1 (ver definição de posto na página 117). Em particular, a multiplicação de matriz por vetor ganha uma interpretação bastante útil:

$$Ax = \sum_{k=1}^{p} A(:,k) \cdot x(k) = x(1)A(:,1) + \ldots + x(p)A(:,p).$$

Aqui o resultado deve ser entendido como uma combinação linear das colunas da matriz A.

A multiplicação de um vetor à esquerda da matriz:

$$x^T A = \sum_{k=1}^{p} A(k,:) \cdot x(k) = x(1)A(1,:) + \ldots + x(p)A(p,:).$$

Nesse caso, o produto vetor por matriz representa uma combinação linear das linhas de A ou, ainda, das colunas de A^T.

A.1.3 Multiplicação: matrizes em blocos

Exemplificaremos algumas decomposições de A e B em blocos que permitam as operações de multiplicação serem efetuadas. Representamos um bloco de uma matriz A por A_{ij}, compreendido como um bloco na linha i e coluna j, em relação a uma estrutura de blocos. As operações em blocos são válidas, desde que as dimensões dos blocos sejam consistentes. No primeiro exemplo, uma matriz em blocos 1×2 multiplica uma outra com blocos 2×1:

$$\begin{pmatrix} A_{11} & A_{12} \end{pmatrix} \cdot \begin{pmatrix} B_{11} \\ B_{21} \end{pmatrix} \quad = \quad \begin{pmatrix} A_{11} \cdot B_{11} + A_{12} \cdot B_{21} \end{pmatrix} \quad = \quad C.$$

Em outro exemplo, uma matriz em blocos 2×1 multiplica uma outra em blocos 1×2:

$$\begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix} \cdot \begin{pmatrix} B_{11} & B_{21} \end{pmatrix} = \begin{pmatrix} A_{11} \cdot B_{11} & A_{11} \cdot B_{21} \\ A_{21} \cdot B_{11} & A_{21} \cdot B_{21} \end{pmatrix} =$$

$$= \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = C$$

A.2 Algumas Matrizes

Seja $A \in \mathbb{C}^{m \times p}$.

- A **transposta** de A, denotada por A^T é tal que $A^T \in \mathbb{C}^{p \times m}$ e $A^T(i,j) = A(j,i)$ para $1 \leq i \leq p$ e $1 \leq j \leq m$, ou seja trocamos as linhas pelas colunas de A.
- A **transposta conjugada** de A, denotada por A^H é tal que $A^H \in \mathbb{C}^{p \times m}$ e $A^H(i,j) = \overline{A(j,i)}$ para $1 \leq i \leq p$ e $1 \leq j \leq m$, ou seja trocamos as linhas pelas colunas de A e substituímos cada valor complexo de A pelo seu conjugado. Usa-se também a notação A^*.

Se A é real então $A^T = A^H$. Utilizaremos constantemente as seguintes matrizes quadradas[1], ou seja, $A \in \mathbb{C}^{m \times m}$.

[1] O conceito que se encontra ao lado esquerdo do símbolo := é definido pela sentença em seu lado direito.

1. A é **hermitiana** $:= A^H = A$.
2. A é **simétrica** $:= A^T = A$.
3. A é **positivo-definida** $:= x \neq 0 \Rightarrow x^H Ax > 0$.
4. A é **positivo-semidefinida** $:= x \neq 0 \Rightarrow x^H Ax \geq 0$.
5. A é **unitária** $:= A^H A = AA^H = I$, onde I é a matriz identidade de ordem m.
6. A é **ortogonal** $:= A^T A = AA^T = I$.
7. A é **normal** $:= A^H A = AA^H$.
8. A é **triangular superior** := se $i > j \Rightarrow A(i,j) = 0$, ou seja, a matriz é necessariamente nula abaixo da diagonal principal.
9. A é **estritamente triangular superior** := se $i \geq j \Rightarrow A(i,j) = 0$, ou seja, a matriz é necessariamente nula abaixo da primeira sobrediagonal.
10. A é **triangular inferior** := se $i < j \Rightarrow A(i,j) = 0$, ou seja, a matriz é necessariamente nula acima da diagonal principal.
11. A é **estritamente triangular inferior** := se $i \leq j \Rightarrow A(i,j) = 0$, ou seja, a matriz é necessariamente nula acima da primeira subdiagonal.
12. A é **diagonal** := A é triangular superior e triangular inferior concomitantemente. É usual a notação $\text{diag}(A(1,1), A(2,2), \ldots A(m,m))$.
13. A é **Hessenberg superior** := se $i > j + 1 \Rightarrow A(i,j) = 0$, ou seja, todas as entradas abaixo da primeira subdiagonal são nulas.

$$\begin{pmatrix} A(1,1) & A(1,2) & \cdots & \cdots & \cdots & A(1,m) \\ A(2,1) & A(2,2) & \cdots & \cdots & \cdots & A(2,m) \\ 0 & A(3,2) & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & A(m,m-1) & A(m,m) \end{pmatrix}$$

14. A é **Hessenberg inferior** := se $i < j+1 \Rightarrow A(i,j) = 0$, ou seja, todas as entradas acima da primeira sobrediagonal são nulas, ou ainda, A^T é Hessenberg superior

15. A é **tridiagonal** := A é Hessenberg superior e A é Hessenberg inferior, concomitantemente.

16. A é **regular** := A tem inversa.

17. A é **singular** := A não tem inversa.

18. A é **diagonal por blocos** := A é da forma

$$\begin{pmatrix} A_{11} & 0 & \dots & 0 \\ 0 & A_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{kk} \end{pmatrix}$$

onde cada bloco $A_{ii} \in \mathbb{C}^{m_i \times m_i}$, $i = 1:k$, é uma matriz quadrada e todos os elementos fora desses blocos são nulos.

Matrizes retangulares herdam algumas dessas mesmas definições, desde que garantida a coerência das sentenças.

A.3 Espaços Relevantes

Seja $A \in \mathbb{C}^{m \times p}$. Utilizaremos a seguinte nomenclatura para nos referirmos a espaços vetoriais e a propriedades ligados a A.

1. O **núcleo** de A: $\text{Nuc}(A) := \{x \in \mathbb{C}^p;\ Ax = 0\}$.

2. A **imagem** de A: $\text{Im}(A) := \{y \in \mathbb{C}^m;\ \text{existe } x \in \mathbb{C}^p;\ Ax = y\}$. Como esse espaço é formado por combinações lineares de colunas de A, utilizaremos, também, o nome **espaço coluna** para nos referirmos a ele.

3. A imagem de A^T será denominada **espaço linha** de A, uma vez que é formada por combinações lineares das linhas de A. A imagem de A^H será denominada **espaço conjugado linha** de A,

uma vez que é formada por combinações lineares dos conjugados das linhas de A.

4. O **posto coluna** é a dimensão do espaço coluna.
5. O **posto linha** é a dimensão do espaço linha ou do espaço conjugado linha.

A.4 Posto

Teorema A.1 (Teorema do Núcleo e da Imagem)**.** *Sejam $\mathcal{V}$ e $\mathcal{W}$ espaços vetoriais e $L : \mathcal{V} \to \mathcal{W}$ uma transformação linear. Sejam* $\dim(\mathcal{V}) = n$, $\dim(\text{Nuc}(L)) = r$, $\dim(\text{Im}(L)) = s$. *Então $n = r + s$, ou seja*

$$\dim(\mathcal{V}) = \dim(\text{Nuc}(L)) + \dim(\text{Im}(L)).$$

Demonstração: Exercício 1.

Esse teorema será complementado, na seção A.5, com o teorema fundamental da álgebra linear.

Teorema A.2 (Teorema do Complemento Ortogonal)**.** *Seja $\mathcal{V}$ um espaço vetorial com* $\dim(\mathcal{V}) = n$ *onde está definido um produto interno. Seja $\mathcal{W}$ um subespaço de $\mathcal{V}$ com* $\dim(\mathcal{W}) = r$. *Seja $\mathcal{W}^\perp$ um subespaço de $\mathcal{V}$ formado por todos os vetores ortogonais a $\mathcal{W}$, denominado* ***complemento ortogonal*** *de $\mathcal{W}$. Então $\mathcal{V}$ é a soma direta de $\mathcal{W}$ e $\mathcal{W}^\perp$*

$$\mathcal{V} = \mathcal{W} \oplus \mathcal{W}^\perp$$

e

$$\dim(\mathcal{V}) = \dim(\mathcal{W}) + \dim(\mathcal{W}^\perp).$$

Demonstração: Exercício 3

Teorema A.3 (Teorema do Posto [76, Teorema 3.1, pág. 166])**.** *Seja A uma matriz $m \times n$. Então o posto linha e o posto coluna de A são iguais a um número r. Além disso* $\dim(\text{Nuc}(A)) = n - r$.

Demonstração: O núcleo da matriz A é formado pelas soluções do sistema homogêneo

$$Ax = 0. \tag{A.4.1}$$

Podemos escrever esse sistema como

$$\sum_{i=1}^{n} x(i)A(:,i) = 0,$$

ou seja o espaço de soluções desse sistema é o núcleo de A e, logo, ambos têm a mesma dimensão. Como o posto coluna de A é igual a dimensão da imagem da transformação linear a qual A está associada, podemos reescrever o teorema do núcleo e da imagem, A.1, como

posto coluna + dimensão do espaço de soluções do sistema A.4.1 $= n$.

Uma outra interpretação do sistema homogêneo (A.4.1) é que as soluções pertencem ao complemento ortogonal do espaço linha de A, uma vez que cada solução é ortogonal a todas as linhas de A, logo, usando o teorema do complemento ortogonal, A.2, temos que

posto linha + dimensão do espaço de soluções do sistema A.4.1 $= n$.

Com as duas igualdades anteriores, as afirmações do enunciado do teorema ficam provadas. ■

O número r calculado no teorema do posto A.3 é chamado **posto** da matriz A, sendo igual ao posto linha e ao posto coluna de A.

A.5 Teorema Fundamental da Álgebra Linear

Teorema A.4 (Teorema Fundamental da Álgebra Linear[2] [122, seções 2.4 e 3.1])**.** *Sejam $A \in \mathbb{C}^{m\times n}$ e r o posto de A.*

1. $\dim(\mathrm{Im}(A)) = \mathrm{postocoluna}(A) = r$.
2. $\dim(\mathrm{Im}(A^H)) = \mathrm{postolinha}(A) = r$.

[2]Estamos usando a nomenclatura proposta por G. Strang em [122], no entanto com uma adaptação para o corpo dos complexos.

3. $\dim(\text{Nuc}(A)) = n - r$.

4. $\dim(\text{Nuc}(A^H)) = m - r$.

5. $\text{Im}(A) = (\text{Nuc}(A^H))^\perp \Rightarrow \mathbb{C}^m = \text{Im}(A) \oplus \text{Nuc}(A^H)$.

6. $\text{Im}(A^H) = (\text{Nuc}(A))^\perp \Rightarrow \mathbb{C}^n = \text{Im}(A^H) \oplus \text{Nuc}(A)$.

A demonstração do, assim chamado, teorema fundamental da álgebra linear utiliza reiteradamente os teoremas A.1, A.2, A.3, e ficará como o exercício 4, no fim desse apêndice.

A.6 Projeções

Seja $\mathcal{V}$ um espaço vetorial. Uma transformação linear P de $\mathcal{V}$ em $\mathcal{V}$ é uma projeção quando $P^2 = P$. Caso $P = P^H$, então P é uma projeção ortogonal, as demais projeções são denominadas oblíquas. A representação matricial dessa transformação linear será discutida no teorema A.6 na página 119.

Propriedade A.1 (Projeções [24, adaptado da seção 4.8.2])**.** *Sejam $\mathcal{V}$ um espaço vetorial e P uma projeção de $\mathcal{V}$ em $\mathcal{V}$. Valem as seguintes propriedades:*

1. *$(I - P)$ é uma projeção,*

2. *P^H é uma projeção,*

3. $\text{Im}(P) = \text{Nuc}(I - P)$,

4. $\text{Im}(I - P) = \text{Nuc}(P)$,

5. $\mathcal{V} = \text{Im}(I - P) \oplus \text{Im}(P^H)$ *e* $\text{Im}(I - P) \perp \text{Im}(P^H)$,

6. $\mathcal{V} = \text{Im}(I - P^H) \oplus \text{Im}(P)$ *e* $\text{Im}(I - P^H) \perp \text{Im}(P)$,

7. *se U é uma matriz cujas colunas são ortonormais então UU^H é uma matriz que representa uma projeção ortogonal em* $\text{Im}(U)$.

Demonstração: Exercício 5.

Uma projeção oblíqua P é, no entanto, ortogonal a um dado subespaço vetorial. Pelas propriedades enunciadas em A.1, $\mathrm{Im}(I-P) = \mathrm{Nuc}(P)$ e $\mathrm{Im}(I-P) \perp \mathrm{Im}(P^H)$ como a projeção é paralela ao subespaço $\mathrm{Im}(I-P)$, ela se dará ortogonalmente ao subespaço $\mathrm{Im}(P^H)$ e nos podemos enunciar o teorema seguinte.

Teorema A.5 ([24, teorema 4.10, pág. 130])**.** *Sejam $\mathcal{V}$ um espaço vetorial e P uma projeção de $\mathcal{V}$ em $\mathcal{V}$.*

$$Py = 0 \Leftrightarrow y \in (\mathrm{Im}(P^H))^{\perp}.$$

Demonstração: Exercício 6.

Teorema A.6 ([24, teorema 4.11, pág. 131])**.** *Sejam P uma projeção, U uma matriz cujas colunas formam uma base ortonormal para* $\mathrm{Im}(P)$ *e V uma matriz cujas colunas formam uma base ortonormal para* $\mathrm{Im}(P^H)$*. Então*

$$(V^H U)$$

é regular e uma matriz que representa a projeção P é dada por

$$U(V^H U)^{-1} V^H.$$

A que representa P^H, por sua transposta conjugada.

Demonstração: Provando a regularidade de $(V^H U)$. Pelo teorema A.4, $\mathrm{Im}(P)$ e $\mathrm{Im}(P^H)$ têm a mesma dimensão, logo $V^H U$ é uma matriz quadrada de ordem, digamos n, menor ou igual à dimensão do espaço vetorial onde P está definida. Caso $V^H U$ não tenha posto completo, existe uma combinação linear de suas linhas cujo resultado é o vetor nulo,

$$\sum_{i=1}^{n} \alpha_i V^H U(i,:) = 0, \tag{A.6.2}$$

onde nem todos os escalares α_i são nulos. Seja j o índice de um dos escalares diferentes de zero. Seja E_j uma matriz onde, a exceção da

$j-$ésima, todas as linhas são iguais às da identidade de ordem n, e a $j-$ésima seja composta pelos n escalares da combinação linear em (A.6.2), tal que o elemento α_i fique na coluna i de E_j. Ora, temos que a matriz resultante da multiplicação $E_j(V^H U)$ tem a sua $j-$ésima linha nula. Consideremos agora a matriz $(E_j V^H)$; a sua $j-$ésima linha é, por hipótese, diferente de zero, pois a linhas de V^H são linearmente independentes, ou seja, essa $j-$ésima linha é um vetor pertencente ao subespaço $\mathrm{Im}(P^H)$ e é, como vimos, diferente de zero. Por outro lado, esse vetor é ortogonal ao espaço $\mathrm{Im}(P)$, pois é ortogonal a todos vetores de uma de suas bases. No entanto, pelo teorema A.4, $\mathrm{Im}(P^H) \perp \mathrm{Nuc}(P)$, logo o único vetor de $\mathrm{Im}(P^H)$ ortogonal a $\mathrm{Im}(P)$ é o vetor nulo, levando a uma contradição. Logo $V^H U$ tem posto completo e, sendo quadrada, tem inversa. Quanto à construção da matriz, temos:

$$y = y - Py + Py = (I - P)y + Py.$$

Sejam VV^H uma projeção ortogonal em $\mathrm{Im}(P^H)$, UU^H uma projeção ortogonal em $\mathrm{Im}(P)$ e $x = Py$, então

$$VV^H y = VV^H((I - P)y + x),$$

como, pelo teorema A.5, $\mathrm{Im}(I-P) = \mathrm{Nuc}(P^H)$, logo $VV^H(I-P)y = 0$ e teremos

$$VV^H y = VV^H x,$$

como $x \in \mathrm{Im}(P)$ então $x = UU^H x$, e

$$\begin{aligned} VV^H y = VV^H UU^H x \Rightarrow V^H y = V^H UU^H x \Rightarrow \\ \Rightarrow (V^H U)^{-1} V^H y = U^H x \Rightarrow U(V^H U)^{-1} V^H y = x. \end{aligned}$$

A demonstração para a matriz P^H é deixada como exercício. ■

A.7 Autovalores e Autoespaços

Se $A \in \mathbb{C}^{m\times m}$, $x \in \mathbb{C}^m$ e $\lambda \in \mathbb{C}$, consideramos[3] a equação

$$Ax = \lambda x, \quad x \neq 0. \tag{A.7.3}$$

Se um escalar λ e um vetor não-nulo x são soluções dessa equação, então λ é um **autovalor** de A e x é um **autovetor** de A associado a λ. O conjunto de todos os escalares complexos que são autovalores de A é denominado **espectro** de A, $\sigma(A)$. O **raio espectral** de A é o número real não-negativo $\rho(A) = \max\{|\lambda|; \lambda \in \sigma(A)\}$.

Lembremo-nos que um polinômio é da forma

$$p(t) = a_k t^k + a_{k-1}t^{k-1} + \ldots + a_1 t + a_0 = a_0 + \sum_{i=1}^{k} a_i t^i.$$

No caso de matrizes quadradas e expoentes positivos, há sentido em definir um polinômio matricial

$$p(A) := a_k A^k + a_{k-1}A^{k-1} + \ldots + a_1 A + a_0 I = a_0 I + \sum_{i=1}^{k} a_i A^i.$$

O próximo teorema estabelece relações entre os autovalores e autovetores de A e de $p(A)$

Teorema A.7. *Se $p(*)$ é um dado polinômio, λ um autovalor de $A \in \mathbb{C}^{m\times m}$ e x um autovetor associado a λ. Então $p(\lambda)$ é um autovalor da matriz $p(A)$ e x é um autovetor de $p(A)$ associado a $p(\lambda)$.*

Demonstração: Exercício 10.

O resultado seguinte apresenta uma condição necessária e suficiente para a singularidade de uma matriz.

Teorema A.8. *$A \in \mathbb{C}^{m\times m}$ é singular $\Leftrightarrow 0 \in \sigma(A)$.*

[3]Esta parte do texto é, essencialmente, um resumo do capítulo 1 de [67].

Demonstração: Exercício 11.

A equação (A.7.3) pode ser reescrita como

$$(\lambda I - A)x = 0, \quad x \neq 0.$$

Logo, $\lambda \in \sigma(A)$ se e somente se $(\lambda I - A)$ é uma matriz singular ou seja

$$\det(\lambda I - A) = 0.$$

Entendido como um polinômio formal em t, o **polinômio característico** de $A \in \mathbb{C}^{m\times m}$ é definido por

$$p_A(t) := \det(tI - A).$$

Um teorema fundamental, ligado ao polinômio característico, é estabelecido a seguir.

Teorema A.9. *Seja $A \in \mathbb{C}^{m\times m}$, o polinômio característico $p_A(*)$ tem grau m e o conjunto de suas raízes coincide com $\sigma(A)$.*

Demonstração: Exercício 16.

Observação A.1. *Cada matriz $A \in \mathbb{C}^{m\times m}$, quando definida sobre o corpo dos complexos, tem exatamente m autovalores, contando as multiplicidades.*

Uma definição útil é a de similaridade entre matrizes. Diremos que duas matrizes A e B, ambas em $\mathbb{C}^{m\times m}$, são **similares** se existe $S \in \mathbb{C}^{m\times m}$, S regular, tal que

$$B = S^{-1}AS.$$

Podemos observar que a relação de similaridade é uma relação de equivalência. Uma matriz $A \in \mathbb{C}^{m\times m}$ é denominada **diagonalizável** caso seja similar a uma matriz diagonal.

Teorema A.10. *$A \in \mathbb{C}^{m\times m}$ é diagonalizável $\Leftrightarrow$ existe uma conjunto de m vetores linearmente independentes que são autovetores de A.*

Demonstração: Exercício 17.

Uma outra caracterização de matrizes diagonalizáveis é dada pelo próximo teorema.

Teorema A.11. *Se $A \in \mathbb{C}^{m\times m}$ tem m autovalores distintos então A é diagonalizável.*

Demonstração: Exercício 19.

Seja $A \in \mathbb{C}^{m\times m}$. Para um dado $\lambda \in \sigma(A)$, o conjunto de todos os vetores $x \in \mathbb{C}^m$, incluindo o vetor nulo, que satisfaçam a equação $Ax = \lambda x$ é chamado de **autoespaço** de A em relação ao autovalor λ. Observe que todos os vetores de um autoespaço, a exceção do vetor nulo, são também autovetores de A em relação ao autovalor λ. A dimensão de um autoespaço de A em relação a um autovalor λ é chamada de **multiplicidade geométrica** do autovalor λ. A multiplicidade de um autovalor λ enquanto raiz do polinômio característico é chamada de **multiplicidade algébrica**.

A.8 Decomposições

A.8.1 Decomposição de Schur

Uma matriz $B \in \mathbb{C}^{m\times m}$ tem a propriedade de ser **equivalente unitariamente** a uma matriz $A \in \mathbb{C}^{m\times m}$, se existe uma matriz unitária $U \in \mathbb{C}^{m\times m}$ tal que $B = U^H AU$. Caso U seja real, então B goza da propriedade de **equivalência ortogonal real** em relação a A.

No teorema a seguir apresentamos uma das decomposições fundamentais da álgebra linear.

Teorema A.12 (Teorema de Schur)**.** *Seja $A \in \mathbb{C}^{m\times m}$ com m autovalores $\lambda_1, \lambda_2, \ldots, \lambda_m$, distintos ou não, e em qualquer ordem dada, então existe uma matriz unitária $U \in \mathbb{C}^{m\times m}$ tal que*

$$U^H AU = T$$

é triangular superior, e cujas entradas diagonais $T(i,i) = \lambda_i,\ i = 1 : m$. *Ou seja, toda matriz quadrada é equivalente unitariamente a uma matriz triangular superior (ou triangular inferior)*

Demonstração: Exercício 20.

Observação A.2. *A decomposição dada no teorema A.12 não é única, mas representa a forma mais simples que se pode colocar uma matriz através de uma equivalência unitária.*

A versão do teorema de Schur, A.12, para matrizes definidas sobre o corpo dos reais é dada pelo teorema abaixo.

Teorema A.13 (Teorema de Schur (versão para reais))**.** *Seja* $A \in \mathbb{R}^{m\times m}$, *então existe uma matriz real ortogonal* $Q \in \mathbb{R}^{m\times m}$ *tal que*

$$Q^T A Q = \begin{pmatrix} A_{11} & * & \dots & * \\ 0 & A_{22} & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{kk} \end{pmatrix}$$

é Hessenberg superior, e cujos blocos diagonais são matrizes reais 1×1 *ou matrizes reais* 2×2, *cujos dois autovalores complexos são pares de números complexos conjugados.*

Demonstração: Exercício 21.

Observação A.3. *A decomposição dada no teorema A.13 representa a forma mais simples que se pode colocar uma matriz através de uma equivalência ortogonal real, pois uma matriz real pode ter autovalores complexos. Observamos ser possível definir um isomorfismo entre o conjunto das matrizes* $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ *e o corpo dos números complexos, onde a soma e a multiplicação entre dois complexos é substituída pela soma e multiplicação entre duas matrizes, essas matrizes tem dois autovalores complexos conjugados* $a + \imath b$ *e* $a - \imath b$.

O teorema de Schur permite uma demonstração simples do próximo teorema que afirma ser a matriz A uma raiz de seu polinômio característico.

Teorema A.14 (Teorema de Cayley-Hamilton)**.** *Seja $A \in \mathbb{C}^{m\times m}$ e $p_A(t)$ seu polinômio característico então*

$$p_A(A) = 0.$$

Demonstração: Exercício 22.

Uma outra aplicação do teorema de Schur é mostrar que toda matriz é "quase diagonalizável", no seguinte sentido:

Teorema A.15. *Seja $A \in \mathbb{C}^{m\times m}$. Para todo $\epsilon > 0$, existe uma matriz $A_\epsilon \in \mathbb{C}^{m\times m}$ que possui m autovalores distintos, logo é diagonalizável, e tal que*

$$\sum_{i,j=1}^{m} |A(i,j) - A_\epsilon(i,j)|^2 < \epsilon.$$

Demonstração: Exercício 24.

O resultado seguinte prepara as condições técnicas para a demonstração do teorema da forma canônica Jordan.

Teorema A.16. *Suponha que $A \in \mathbb{C}^{m\times m}$ possua k autovalores distintos λ_i, com multiplicidades m_i, $i = 1 : k$. Então A é similar a uma matriz diagonal por blocos*

$$T = \begin{pmatrix} T_{11} & 0 & \dots & 0 \\ 0 & T_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T_{kk} \end{pmatrix} \quad \textit{onde} \quad T_{ii} = \begin{pmatrix} \lambda_i & * & \dots & * \\ 0 & \lambda_i & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_i \end{pmatrix}$$

ou seja, cada $T_{ii} \in \mathbb{C}^{m_i\times m_i}$, $i = 1 : k$, é uma matriz triangular superior onde todas as entradas na diagonal principal são iguais ao autovalor λ_i.

Demonstração: Exercício 25.

A.8.2 Forma Canônica de Jordan

Começamos essa seção com citações que informam um pouco sobre as vantagens e limitações da forma canônica de Jordan.

> Many authors have made this theorem the climax of their linear algebra course. Frankly, I think that is a mistake. It is certainly true that not all matrices are diagonalizable, and that the Jordan form is the most general case: but for that very reason its construction is both technical and extremely unstable. (A slight change in A can put back all the missing eigenvectors, and remove the off-diagonal l's.)[122, pág. 312]

> Since the Jordan form of a matrix need not be a continuous function of the entries of the matrix, it is possible that small variations in the entries of a matrix will result in large variations in the entries of the Jordan form. There is no hope of computing such an object in a stable way, so the Jordan canonical form is little used in numerical applications.
>
> Despite this limitation, the Jordan canonical form is well worth knowing and is a rich source of insights. As a matter of general technique, if one has something to prove about matrices it is well to consider first if it can be proved for diagonal matrices and, if this is successful, then to see if some limiting argument may establish the result in general (using the fact that any complex matrix can be approximated arbitrarily closely by a diagonalizable matrix). If this does not work, or if one prefers to avoid an analytical argument, one might next try to prove the result for upper triangular or Jordan matrices. It is sometimes useful to know that every matrix is similar to a matrix of the form (3.1.12)[4] in which all the "+ 1"terms in the Jordan blocks are replaced by $\epsilon > 0$ and ϵ can be taken to be arbitrarily small. [67, pág. 128]

> However, as a mathematical probe the Jordan canonical

[4] Trata-se do teorema A.15 no presente texto.

> form is still useful, and reports of its death are greatly exaggerated. [120, pág. 22]

As citações encontram respaldo no exercício 26. Advertências feitas, passamos à apresentação dessa decomposição, começando por algumas definições. Um **bloco de Jordan** $J_k(\lambda)$ é uma matriz quadrada de ordem k, triangular superior, da forma

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ 0 & 0 & \lambda & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \dots & \lambda \end{pmatrix}$$

Há $k-1$ termos "$+1$" que aparecem na primeira sobrediagonal e o escalar λ encontra-se nas k posições da diagonal principal do bloco; todas as demais entradas são nulas. Uma **matriz de Jordan** $J \in \mathbb{C}^{m\times m}$ é a soma direta de blocos de Jordan

$$J = \begin{pmatrix} J_{m_1}(\lambda_1) & 0 & 0 & \dots & 0 \\ 0 & J_{m_2}(\lambda_2) & 0 & \dots & 0 \\ 0 & 0 & \lambda & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & J_{m_k}(\lambda_k) \end{pmatrix}, \quad m = \sum_{i=1}^{k} m_i,$$

onde nem as ordens m_i nem os escalares λ_i precisam ser distintos.

Uma das possíveis demonstrações para o teorema da forma canônica de Jordan segue os seguintes passos:

1. Usar o teorema da triangulação de Schur A.12 para transformar a matriz qualquer dada, em uma matriz similar com uma outra na forma triangular superior.

2. Usar o teorema A.16 para transformar a matriz triangular superior em uma matriz diagonal por blocos, onde cada bloco é uma matriz triangular superior.

3. Mostrar que uma matriz triangular superior, cujas entradas da diagonal principal são iguais, é similar a uma soma direta de blocos de Jordan.

Teorema A.17 (Teorema da Forma Canônica de Jordan)**.** *Seja* $A \in \mathbb{C}^{m\times m}$ *então existe uma matriz regular* $S \in \mathbb{C}^{m\times m}$*, tal que*

$$A = S \begin{pmatrix} J_{m_1}(\lambda_1) & 0 & 0 & \dots & 0 \\ 0 & J_{m_2}(\lambda_2) & 0 & \dots & 0 \\ 0 & 0 & \lambda & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & J_{m_k}(\lambda_k) \end{pmatrix} S^{-1} = SJS^{-1},$$

com $m = \sum_{i=1}^{k} m_i$*.*

Demonstração: Estude as obras de referência.

Observação A.4. *Qualquer bloco de Jordan* $J_k(\lambda)$ *pode ser descrito como* $J_k(\lambda) = \lambda I + N_k$ *onde* $(N_k)^k = 0$*, ou seja, tem índice de nilpotência* k*. Generalizando esse fato, qualquer matriz de Jordan pode ser escrita como* $J = D + N$*, onde* D *é uma matriz diagonal, cuja diagonal principal é igual a de* J *e* $N = J - D$*. A matriz* N *é nilpotente e seu índice de nilpotência* k *é igual a ordem do maior bloco de Jordan de* J*.*

Observação A.5. *Seja* $A \in \mathbb{C}^{m\times m}$ *então, pelo teorema da forma canônica de Jordan,* $A = SJS^{-1}$ *então* $A = SDS^{-1} + SNS^{-1}$*, onde a primeira parcela é diagonalizável e a segunda nilpotente. Concluímos que toda matriz pode ser escrita como a soma de uma matriz diagonalizável e uma outra nilpotente.*

Observação A.6. *A multiplicidade geométrica de um autovalor de uma dada matriz* $A \in \mathbb{C}^{m\times m}$ *é igual ao número de blocos de Jordan associados a esse autovalor. Esse número é menor ou igual a soma de todas as ordens dos blocos associados a esse autovalor. Essa última soma é a multiplicidade algébrica desse autovalor.*

Vamos precisar da próxima definição para o enunciado do teorema a seguir. Um **polinômio mônico** tem o coeficiente associado ao termo de mais alta ordem igual a "+1".

Teorema A.18 (Teorema do Polinômio Mínimo)**.** *Seja $A \in \mathbb{C}^{m\times m}$, então existe um único polinômio mônico, $q_A(t)$, de grau mínimo que anula A. O grau desse polinômio é no máximo igual a m. Se $p(t)$ é um polinômio qualquer para o qual $p(A) = 0$ então $q_A(t)$ divide $p(t)$. Esse polinômio é chamado de* ***polinômio mínimo*** *de A.*

Demonstração: Exercício 27.

Teorema A.19. *Seja $A \in \mathbb{C}^{m\times m}$ cujos autovalores distintos são $\lambda_1, \lambda_2, \ldots, \lambda_n$. O polinômio mínimo de A é*

$$q_A(t) = \prod_{i=1}^{n} (t - \lambda_i)^{r_i}$$

onde cada r_i é a ordem do maior bloco de Jordan de A associado ao autovalor λ_i.

Demonstração: Exercício 27.

Esse resultado é utilizado na discussão da conveniência do métodos de Krylov.

A.8.3 Decomposição em Valores Singulares

Essa é considerada uma das decomposições centrais da álgebra linear.

Teorema A.20 (Decomposição em Valores Singulares)**.** *Seja $A \in \mathbb{C}^{m\times n}$ de posto k , então ela pode ser decomposta da seguinte forma.*

$$A = V\Sigma W^H$$

onde $V \in \mathbb{C}^{m\times m}$, $\Sigma \in \mathbb{C}^{m\times n}$ e $W \in \mathbb{C}^{n\times n}$ têm as seguintes propriedades:

1. *V e W são matrizes unitárias,*

2. $\Sigma(i,j) = 0$ *para todo* $i \neq j$,

3. $\Sigma(1,1) \geq \Sigma(2,2) \geq \ldots \geq \Sigma(k,k) > \Sigma(k+1,k+1) = \ldots = \Sigma(q,q) = 0$, *aonde* $q = \min(m,n)$. *É usual a notação:* $\sigma_i := \Sigma(i,i)$,

4. *Os escalares* σ_i *são as raízes quadradas, portanto não-negativos, dos autovalores de* AA^H, *e, por isso, unicamente determinados,*

5. *As colunas de* V *são os autovetores de* AA^H *e as colunas de* W *os autovetores de* $A^H A$.

Demonstração: Estude as obras de referência.

Observação A.7. *Os elementos* σ_i, $i = 1 : q = \min(m,n)$ *da diagonal de* Σ *são denominados* ***valores singulares*** *de* $A \in \mathbb{C}^{m\times n}$. *Alguns autores consideram como valores singulares apenas o elementos positivos, ou seja, somente os* σ_i, $i = 1 : k$.

Observação A.8. *As colunas de* V *são os* ***vetores singulares à esquerda*** *de* A *e as colunas de* W *são os* ***vetores singulares à direita*** *de* A.

Teorema A.21. *Seja* $A \in \mathbb{C}^{m\times m}$ *e* $A = V\Sigma W^H$ *uma decomposição em valores singulares, caso* A *seja regular, então:*

$$A = \sum_{j=1}^{m} \sigma_j v_j w_j^H \quad e \quad A^{-1} = \sum_{j=1}^{m} \frac{1}{\sigma_j} w_j v_j^H$$

Demonstração: Exercício 31.

A.9 Normas de Vetores

Ao analisarmos os métodos de Krylov nos deparamos com a necessidade de medirmos a distância entre uma solução aproximada e a solução exata, ou o comprimento de um vetor resíduo, ou, ainda, o

"tamanho" de uma matriz; para tanto precisamos saber medir vetores e matrizes. Até o final desse apêndice trataremos desses temas.

Seja $\mathcal{V}$ um espaço vetorial definido sobre um corpo $\mathbb{F}$ ($\mathbb{R}$ ou $\mathbb{C}$). A função $\|*\| : \mathcal{V} \to \mathbb{R}$ é uma **norma vetorial** para todo $x, y \in \mathcal{V}$ e para todo $\alpha \in \mathbb{F}$, se atende as seguintes propriedades:

1. $\|x\| \geq 0$ - não-negatividade,
2. $\|x\| = 0 \Leftrightarrow x = 0$ - positividade,
3. $\|\alpha x\| = |\alpha|\|x\|$ - homogeneidade,
4. $\|x + y\| \leq \|x\| + \|y\|$ - desigualdade triangular.

Uma função que atende às propriedades 1, 3 e 4, mas não atende a 2 é chamada de **seminorma vetorial**.

Outra noção necessária é a de ângulo entre vetores (ver exercício 30), para tanto introduzimos o conceito de produto interno, ou produto escalar. Seja $\mathcal{V}$ um espaço vetorial definido sobre um corpo $\mathbb{F}$ ($\mathbb{R}$ ou $\mathbb{C}$). A função $(*,*) : \mathcal{V} \times \mathcal{V} \to \mathbb{F}$ é um **produto interno** para todo $x, y, z \in \mathcal{V}$ e para todo $\alpha \in \mathbb{F}$, se atende as seguintes propriedades:

1. $(x, x) \geq 0$ - não-negatividade,
2. $(x, x) = 0 \Leftrightarrow x = 0$ - positividade,
3. $(x + y, z) = (x, z) + (y, z)$ - aditividade,
4. $(\alpha x, y) = \alpha(x, y)$ - homogeneidade,
5. $(x, y) = \overline{(y, x)}$ - propriedade hermitiana.

Um resultado básico é o seguinte.

Teorema A.22 (Desigualdade de Cauchy-Schwarz)**.** *Se $(*,*)$ é um produto interno definido no espaço vetorial $\mathcal{V}$ sobre um corpo $\mathbb{F}$ ($\mathbb{R}$ ou $\mathbb{C}$) então*

$$|(x, y)|^2 \leq (x, x)(y, y).$$

A igualdade ocorre apenas quando os vetores são linearmente dependentes.

Demonstração: Exercício 29.

Corolário A.1. *Se $(*,*)$ é um produto interno definido no espaço vetorial $\mathcal{V}$, então $\|x\| = (x,x)^{1/2}$ é uma norma vetorial em $\mathcal{V}$. Nesse caso diremos que a norma vetorial é derivada de um produto interno.*

A.9.1 Exemplos de Normas Vetoriais

Apresentamos alguns exemplos usuais de normas vetoriais:

1. A **norma euclidiana** ou **norma $\mathbf{l_2}$**, definida em $\mathbb{C}^m$, é dada por

$$\|x\|_2 := (|x_1|^2 + |x_2|^2 + \ldots + |x_m|^2)^{1/2}.$$

   Esta norma é derivada do **produto interno euclidiano**, ou seja, $\|x\|_2^2 = (x,x) = x^H x$.

2. A **norma da soma** ou **norma $\mathbf{l_1}$**, definida em $\mathbb{C}^m$, é dada por

$$\|x\|_1 := |x_1| + |x_2| + \ldots + |x_m|.$$

   Essa norma não é derivada de um produto interno.

3. A **norma do máximo** ou **norma infinito**, definida em $\mathbb{C}^m$, é dada por

$$\|x\|_\infty := \max\{|x_1|, |x_2|, \ldots, |x_m|\}.$$

4. A norma $\mathbf{l_p}$, definida em $\mathbb{C}^m$, é dada por

$$\|x\|_p = \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p}$$

   para $p \geq 1$.

Um resultado que será muito utilizado na discussão dos métodos de Krylov, relaciona a norma euclidiana com os operadores de projeção ortogonal.

Teorema A.23. *Let $\mathcal{V} \subset \mathbb{C}^m$ um subespaço e $z \in \mathbb{C}^m$. Então a solução do problema*

$$\arg\min_{x\in\mathcal{V}} \|z - x\|_2$$

é a projeção ortogonal de z em $\mathcal{V}$ e esse mínimo

$$\min_{x\in\mathcal{V}} \|z - x\|_2$$

vale $\|P_\perp z\|_2$, onde $P_\perp$ é a projeção ortogonal no complemento ortogonal de $\mathcal{V}$.

Demonstração: Essa demonstração tem várias versões, vamos apresentar os detalhes de uma delas. Seja V uma matriz cujas colunas são formada por uma base ortogonal de $\mathcal{V}$. Logo $P := VV^H$ é uma projeção ortogonal em $\mathcal{V}$ e $P_\perp := (I - VV^H)$ é uma projeção ortogonal no complemento ortogonal de $\mathcal{V}$. Temos que

$$z-x = z-x+P(z-x)-P(z-x) = P(z-x)+P_\perp(z-x) = Pz-x+P_\perp z,$$

logo, podemos resolver o problema equivalente

$$\min_{x\in\mathcal{V}} \|Pz - x + P_\perp z\|_2$$

desenvolvendo a norma como um produto interno, temos

$$\begin{aligned}\|Pz - x + P_\perp z\|_2 &= \big(Pz - x + P_\perp z\ ,\ Pz - x + P_\perp z\big)^{1/2} = \\ &= \big((Pz, Pz) - (x, Pz) - (Pz, x) + (x, x) + (P_\perp z, P_\perp z)\big)^{1/2} = \\ &= \big((\|Pz - x\|_2^2 + \|P_\perp z\|_2^2\big)^{1/2}.\end{aligned}$$

Dado um z qualquer, a única forma de diminuir o valor dessa fórmula é em $\|Pz - x\|_2$, como é uma norma, o mínimo é igual a zero, o que só ocorre quando $x = Pz$, o que implica em

$$\min_{x\in\mathcal{V}} \|z - x\|_2 = \|P_\perp z\|_2.$$

■

A.10 Normas de Matrizes

O conjunto de todas as matrizes $A \in \mathbb{C}^{m\times m}$ é, ele próprio um, espaço vetorial de dimensão m^2, logo, poder-se-ia medir matrizes usando as normas vetoriais em $\mathbb{C}^{m^2}$. Inclusive uma norma utilizada constantemente em álgebra linear computacional é a **norma de Frobenius**: $\|A\|_F = \left(\sum_{j=1}^{m}\sum_{i=1}^{m}(A(i,j))^2\right)^{1/2}$. No entanto, a existência da multiplicação entre matrizes e suas diversas interpretações em vários áreas da matemática induzem a criação de uma medida para essa operação e para as matrizes.

Sejam $\mathcal{M}_m$ o espaço vetorial formado por todas as matrizes $A \in \mathbb{C}^{m\times m}$, A, $B \in \mathcal{M}_m$ e $\alpha \in \mathbb{C}$. A função $\|*\| : \mathcal{M}_m \to \mathbb{R}$ é uma norma de matriz[5] caso atenda as seguintes propriedades[6]:

1. $\|A\| \geq 0$ - não-negatividade,
2. $\|A\| = 0 \Leftrightarrow A = 0$ - positividade,
3. $\|\alpha A\| = |\alpha|\ \|A\|$ - homogeneidade,
4. $\|A + B\| \leq \|A\| + \|B\|$ - desigualdade triangular,
5. $\|AB\| \leq \|A\|\ \|B\|$ - submultiplicatividade.

A.11 Norma Induzida e Raio Espectral

As normas mais usuais em álgebra linear computacional são normas que são definidas a partir de normas vetoriais e, por isso, são denominadas **normas induzidas**. Chama-se norma de matriz induzida pela norma vetorial $\|*\|$ a seguinte função $\|*\| : \mathcal{M}_m \to \mathbb{R}$:

$$\|A\| := \max_{\|x\|=1} \|Ax\|.$$

[5] Alguns autores, incluindo [67], usam a notação $|||*|||$ para denotar uma norma de matriz.

[6] Há autores, por exemplo [94], que usam um versão reduzida de axiomas para a caracterização de normas matriciais. Nesse trabalho estamos apresentando esse conceito apenas para matrizes quadradas, mas ele pode ser estendido sem maiores problemas para matrizes retangulares.

Teorema A.24. *Sejam as seguintes normas induzidas* $\|A\|_1$, $\|A\|_2$ *e* $\|A\|_\infty$, *relativas às seguintes normas vetoriais* $\|x\|_1$, $\|x\|_2$ *e* $\|x\|_\infty$, *respectivamente. Então*

1. $\|A\|_1 = \max_{1\leq j\leq m} \sum_{i=1}^{m} |A(i,j)|$, *ou seja, o máximo da soma dos valores absolutos dos elementos das colunas de* A.
2. $\|A\|_2 = \max\{\sqrt{\lambda}; \lambda \in \sigma(A^H A)\} = \sigma_1(A)$, *o maior valor singular de* A. *Se* A *é hermitiana então* $\|A\|_2 = \rho(A)$.
3. $\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{m} |A(i,j)|$, *ou seja, o máximo da soma dos valores absolutos dos elementos das linhas de* A.

Demonstração: Exercício 34.

A seguir enunciamos algumas propriedades que relacionam normas matriciais e o raio espectral de uma matriz.

Teorema A.25. *Sejam* $A \in \mathbb{C}^{m\times m}$ *e* $\rho(A)$ *seu raio espectral, então temos as seguintes propriedades:*

1. *Para toda norma, induzida ou não*
$$\rho(A) \leq \|A\|.$$
2. *Se* A *é diagonalizável, existe uma norma induzida (dependente de* A*) tal que*
$$\rho(A) = \|A\|.$$
3. *(Householder-Ostrowski) Para toda* A *e para todo* $\epsilon > 0$, *existe ao menos uma norma induzida (dependente de* A *e de* ϵ*), tal que*
$$\|A\| \leq \rho(A) + \epsilon.$$
4. *As condições seguintes são equivalentes:*
   (a) $\lim_{k\to\infty} A^k = 0$,
   (b) $\lim_{k\to\infty} A^k x = 0$, *para todo* $x \in \mathbb{C}^m$,
   (c) $\rho(A) < 1$,

(d) $\|A\| < 1$ *para ao menos uma norma induzida* $\|*\|$.

5. *Uma condição suficiente para que uma matriz* $I-A$ *seja inversível é que* $\rho(A) < 1$, *nesse caso*

$$(I-A)^{-1} = \sum_{j=1}^{\infty} A^j.$$

6. *Para toda* A *e toda a norma matricial, induzida ou não, temos que*

$$\lim_{j\to\infty} \|A^j\|^{1/j} = \rho(A).$$

Demonstração: Exercício 35.

O próximo conceito é ferramenta essencial na análise de métodos de solução de sistemas lineares e será utilizado no teorema a seguir. Chama-se **condicionamento** da matriz regular A o número[7]

$$\kappa_p(A) = \|A\|_p \|A^{-1}\|_p.$$

Teorema A.26. *Seja* $A \in \mathbb{C}^{m\times m}$, *matriz regular,* σ_1 *e* σ_m *o maior e menor valores singulares de* A, *respectivamente, então valem as seguintes propriedades:*

1. $\|A\|_2 = \sigma_1$,
2. $\|A^{-1}\|_2 = \frac{1}{\sigma_m}$,
3. $\kappa_2(A) = \frac{\sigma_1}{\sigma_m}$,
4. $\sigma_m \leq \frac{\|Ax\|_2}{\|x\|_2} \leq \sigma_1,\ \forall x \neq 0.$

Demonstração: Exercício 36.

[7]Esse conceito pode ser estendido para matrizes retangulares usando-se a inversa generalizada, ver [65, pág. 382].

Exercícios

1. Demonstre o teorema A.1.

2. Dada uma base U para um subespaço vetorial $\mathcal{U} \subset \mathbb{C}^m$ sobre o corpo dos números complexos. Seja um conjunto, $\overline{U}$, formado pelos conjugados dos vetores que formam a base U. $\overline{U}$ forma uma base para algum subespaço vetorial de $\mathbb{C}^m$? Se formar, os dois conjuntos são base para um mesmo subespaço? Discuta condições, se for o caso, em que essas afirmações são verdadeiras.

3. Demonstre o teorema A.2.

4. Demonstre o teorema A.4.

5. Demonstre as propriedades apresentadas em A.1.

6. Demonstre o teorema A.5.

7. (Baseado em [66, pág. 311]) Sejam $\mathcal{R}$ e $\mathcal{N}$ subespaços de $\mathcal{V}$, tais que $\mathcal{V} = \mathcal{R} \oplus \mathcal{N}$, então existe uma e apenas uma projeção P tal que $\mathrm{Im}(P) = \mathcal{R}$ e $\mathrm{Nuc}(P) = \mathcal{N}$.

8. Refaça a parte da demonstração do teorema A.6 referente à regularidade da matriz $V^H U$, utilizando uma matriz E_j que multiplique à direita de $V^H U$.

9. Construa a matriz para a projeção P^H, como descrita no teorema A.6.

10. Demonstre o teorema A.7.

11. Demonstre o teorema A.8.

12. ([67, pág. 37]) Considere a matriz diagonal por blocos

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}, \quad A_{ii} \in \mathbb{C}^{m_i \times m_i}.$$

Mostre que $\sigma(A) = \sigma(A_{11}) \cup \sigma(A_{22})$.

13. ([67, pág. 37]) $A \in \mathbb{C}^{m\times m}$ é chamada **idempotente** caso $A^2 = A$. Mostre que cada autovalor de uma matriz idempotente é 0 ou 1.

14. ([67, pág. 37]) $A \in \mathbb{C}^{m\times m}$ é chamada **nilpotente** caso $A^q = 0$, para algum inteiro positivo q. O menor q é denominado **índice de nilpotência**. Mostre que todo autovalor de uma matriz nilpotente é 0.

15. Mostre que todos os autovalores de uma matriz hermitiana são reais.

16. Demonstre o teorema A.9.

17. Demonstre o teorema A.10.

18. [67, pág. 47]) Mostre que a matriz $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ não é diagonalizável.

19. Demonstre o teorema A.11.

20. Demonstre o teorema A.12.

21. Demonstre o teorema A.13.

22. Demonstre o teorema A.14.

23. [67, pág. 87]) Qual o erro no seguinte argumento para justificar a sentença $p_A(A) = 0$? "Como $p_A(\lambda) = 0$ para todos os autovalores λ de $A \in \mathbb{C}^{m\times m}$ e como os autovalores de $q(A)$, onde q é um polinômio, são os escalares $q(\lambda)$ segue que todos os autovalores de $p_A(A)$ são nulos, logo $p_A(A) = 0$". Dê um contra-exemplo para a afirmação.

24. Demonstre o teorema A.15.

25. Demonstre o teorema A.16.

26. (baseado em [67, pág. 127]) Calcule, utilizando a função *jordan* do Matlab, a forma canônica de Jordan da matriz $\begin{pmatrix} eps & 0 \\ 1 & 0 \end{pmatrix}$, onde *eps* é uma constante do Matlab. Observe o resultado e faça

os cálculos por sua própria conta. Reflita sobre a estabilidade do cálculo da forma canônica de Jordan quando $eps \to 0$.

27. Demonstre o teorema A.18.

28. [67, pág. 421]) Seja $A \in \mathbb{C}^{m\times m}$ com sua decomposição em valores singulares dada por $A = V\Sigma W^H$ e defina $A^\dagger = W\Sigma^\dagger V^H$, onde $\Sigma^\dagger$ é a transposta de Σ com os valores singulares de A sendo substituídos por seus inversos multiplicativos. Mostre que:

    (a) $AA^\dagger$ e $A^\dagger A$ são hermitianas,

    (b) $AA^\dagger A = A$ e

    (c) $A^\dagger AA^\dagger = A^\dagger$.

29. Demonstre o teorema A.22.

30. [67, pág. 262]) Mostre que se definirmos ângulo entre dois vetores não-nulos como sendo o valor de

$$\cos^{-1}\left(\frac{|(x,y)|}{((x,x)(y,y))^{1/2}}\right)$$

    que se encontra entre 0 e $\pi/2$, então o conceito de **ângulo** é bem definido para qualquer produto interno.

31. Demonstre o teorema A.21.

32. Verifique que a norma vetorial euclidiana é invariante unitariamente, ou seja, se U é uma matriz unitária, então $\|Ux\|_2 = \|x\|_2$. As norma l_1 e l_∞ também o são?

33. [67, pág. 265]) A norma vetorial do máximo é proveniente de um produto interno?

34. Demonstre o teorema A.24.

35. Demonstre as várias propriedades enunciadas no teorema A.25.

36. Demonstre o teorema A.26.

Bibliografia

[1] G. Allaire and S. M. Kaber. *Numerical Linear Algebra*. Springer, 2008.

[2] P. Arbenz, O. Chinellato, M. Sala, P. Arbenz, and M. H. Gutknecht. *Software for numerical linear algebra*. ETH, 2006.

[3] M. Arioli, I. Duff, and D. Ruiz. Stopping criteria for iterative solvers. *SIAM J. Matrix Anal. Appl.*, 13(1):138–144, January 1992.

[4] M. Arioli, I. S. Duff, S. Gratton, and S. Pralet. A note on GMRES preconditioned by a perturbed LDL^T decomposition with static pivoting. *SIAM Journal on Scientific Computing*, 29(5):2024–2044, 2007.

[5] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quart. of Appl. Math.*, 9(1):17–29, 1951.

[6] O. Axelsson and P. S. Vassilevski. A black box generalized conjugate gradient solver with inner iterations and variable-step preconditioning. *SIAM Journal on Matrix Analysis and Applications*, 12(4):625–644, 1991.

[7] J. Baglama, D. Calvetti, G. H. Golub, and L. Reichel. Adaptively preconditioned GMRES algorithms. *SIAM Journal on Scientific Computing*, 20(1):243–269, 1998.

[8] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, editors. *Templates for the solution of Algebraic Eigenvalue Problems: A Practical Guide*. SIAM, Philadelphia, 2000.

[9] A. H. Baker, J. M. Dennis, and E. R. Jessup. On improving linear solver performance: A block variant of GMRES. *SIAM Journal on Scientific Computing*, 27(5):1608–1626, 2006.

[10] S. Balay, K. Buschelman, V. Eijkhout, W. Gropp, D. Kaushik, M. Knepley, L. C. McInnes, B. Smith, and H. Zhang. PETSc user's manual. Technical Report ANL-95/11, Argonne National Laboratory, 2005. Revision 2.3.0.

[11] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM press, Philadelphia, 1994.

[12] C. Beattie. Harmonic Ritz and Lehman bounds. *ETNA*, 7:18–39, 1998.

[13] M. Benzi. Preconditioning techniques for large linear systems: a survey. *Journal of Computational Physics*, 182(2):418–477, November 2002.

[14] M. Benzi, J. K. Cullum, and M. Tůma. Robust approximate inverse preconditioning for the conjugate gradient method. *SIAM Journal on Scientific Computing*, 22:1318–1332, 2000.

[15] M. Benzi, R. Kouhia, and M. Tůma. Stabilized and block approximate inverse preconditioners for problems in solid and structural mechanics. *Computer methods in applied mechanics and engineering*, 190:6533–6554, 2001.

[16] M. Benzi, C. D. Meyer, and M. Tůma. A sparse approximate inverse preconditioner for the conjugate gradient method. *SIAM Journal on Scientific Computing*, 17:1135–1149., 1996.

[17] M. Benzi and M. Tůma. A comparative study of sparse approximate inverse preconditioners. *Applied Numerical Mathematics*, 30:305–340, 1999.

[18] A. Bouras and V. Frayssé. A relaxation strategy for inexact matrix-vector products for Krylov methods. Technical Report TR/PA/00/15, CERFACS, Toulouse, France, 2000. Submitted to SIAM Journal in Matrix Analysis and Applications.

[19] A. Bouras and V. Frayssé. A relaxation strategy for the Arnoldi method in eigenproblems. Technical Report TR/PA/00/16, CERFACS, Toulouse, France, 2000.

[20] A. Bouras and V. Frayssé. Inexact matrix-vector products in Krylov methods for solving linear systems: a relaxation strategy. *SIAM Journal on Matrix Analysis and Applications*, 26(3):660–678, 2005.

[21] A. Bouras, V. Frayssé, and L. Giraud. A relaxation strategy for inner-outer linear solvers in domain decomposition methods. Technical Report TR/PA/00/17, CERFACS, Toulouse, France, 2000.

[22] R. Bouyouli, K. Jbilou, R. Sadaka, and H. Sadok. Convergence properties of some block Krylov subspace methods for multiple linear systems. *J. Comput. Appl. Math.*, 196(2):498–511, 2006.

[23] M. Brezina, A. J. Cleary, R. D. Falgout, V. E. Henson, J. E. Jones, T. A. Manteuffel, S. F. McCormick, and J. W. Ruge. Algebraic multigrid based on element interpolation (AMGe). *SIAM Journal on Scientific Computing*, 22(5):1570–1592, 2001.

[24] C. Brezinski. *Outils d'analyse numérique pour l'automatique*, chapter Résolution des systèmes linéaires, pages 109–144. Hermes Science Publications, 2002.

[25] C. Brezinski and M. Redivo-Zaglia. Block projection methods for linear systems. *Numerical Algorithms*, 29(1-3):33–43, March 2002.

[26] C. Broyden. Look-ahead block-CG algorithm. In G. W. Althaus and E. Spedicato, editors, *Algorithms for Large Scale Linear Algebraic Systems*. Kluwer Academic,, 1998.

[27] C. L. Calvez and B. Molina. Implicitly restarted and deflated GMRES. *Numerical Algorithms*, 21:261–285, 1999.

[28] L. M. Carvalho. *Preconditioned Schur complement methods in distributed memory environments.* PhD thesis, INPT/CERFACS, France, Oct. 1997. TH/PA/97/41, CERFACS.

[29] L. M. Carvalho, L. Giraud, and P. Le Tallec. Algebraic two-level preconditioners for the Schur complement method. *SIAM J. Scientific Computing*, 22(6):1987–2005, 2001.

[30] L. M. Carvalho, L. Giraud, and G. Meurant. Local preconditioners for two-level nonoverlapping domain decomposition methods. *Numerical Linear Algebra with Applications*, 8(4):207 – 227, 2001.

[31] L. M. Carvalho, S. Gratton, R. Lago, and N, Maculan. Álgebra Linear Numérica e Computacional: Métodos de Krylov para a Solução de Sistemas Lineares. *Livraria Ciência Moderna*, 2010.

[32] F. Chaitin-Chatelin and V. Frayssé. *Lectures on Finite Precision Computations.* SIAM, Philadelphia, 1996.

[33] T. Chan and H. van der Vorst. *Parallel Numerical Algorithms, ICASE/LaRC Interdisciplinary Series in Science and Engineering*, chapter Approximate and incomplete factorizations, pages 167–202. Kluwer, Dordrecht, 1997.

[34] T. Chan and W. Wan. Analysis of projection methods for solving linear systems with several right-hand sides. *SIAM J. Sci. Comput.*, 18(6):1698–1721, 1997.

[35] A. Chapman and Y. Saad. Deflated and augmented Krylov subspace techniques. *Numer. Linear Algebra Appli.*, 4(1):43–66, 1997.

[36] K. Chen. *Matrix Preconditioning Techniques and Applications*. Cambridge University Press, 2005.

[37] E. Chow and Y. Saad. Approximate inverse techniques for block-partitioned matrices. *SIAM Journal on Scientific Computing*, 18(6):1657–1675, 1997.

[38] E. Chow and Y. Saad. Approximate inverse preconditioners via sparse-sparse iterations. *SIAM Journal on Scientific Computing*, 19(3):995–1023, 1998.

[39] J. K. Cullum. Iterative methods for solving $Ax = b$ GMRES-FOM versus QMR/BiCG. Technical Report TR-96-2, Institute for Advances Studies, University of Maryland, 1996.

[40] N. Cundy, J. van den Eshof, A. Frommer, S. Krieg, T. Lippert, and K. Schafer. Numerical methods for the QCD overlap operator iii: Nested iterations. *Comput. Phys. Comm.*, 165:221–242, 2005.

[41] E. de Sturler. Nested Krylov methods based on GCR. *J. Comput. Appl. Math.*, 67(1):15–41, 1996.

[42] E. de Sturler. Truncation strategies for optimal Krylov subspace methods. *SIAM Journal on Numerical Analysis*, 36(3):864–889, 1999.

[43] J. Drkošová, M. Rozložník, Z. Strakoš, and A. Greenbaum. Numerical stability of the GMRES method. *BIT*, 35:309–330, 1995.

[44] M. Eiermann and O. G. Ernst. Geometric aspects of the theory of Krylov subspace methods. *Acta Numerica*, 10:251–312, 2001.

[45] M. Eiermann, O. G. Ernst, and O. Schneider. Analysis of acceleration strategies for restarted minimal residual methods. *Journal of Computational and Applied Mathematics*, 123:261–292, 2000.

[46] M. Embree. How descriptive are GMRES convergence bounds? Technical report, Oxford University Computing Laboratory, 1999.

[47] J. Erhel, K. Burrage, and B. Pohl. Restarted GMRES preconditioned by deflation. *Journal of Computational and Applied Mathematics*, 69(2,):303–318, May 1996.

[48] V. Frayssé, L. Giraud, S. Gratton, and J. Langou. A set of GMRES routines for real and complex arithmetics on high performance computers. *ACM Trans. Math. Softw.*, 31(2):228–238, 2005.

[49] R. Freund and M. Malhotra. A block QMR algorithm for non-Hermitian linear systems with multiple right-hand sides. *Linear Algebra and its Applications*, 254(1–3):119–157, 1997.

[50] K. F. Gauss. *Theory of the Combination of Observations Least Subject to Errors, Part One, Part Two, Supplement.* (Translated by G. W. Stewart) SIAM, 1995.

[51] A. George and J. Liu. *Computer solution of large sparse positive definite systems.* Computational Mathematics. Prentice-Hall, Englewood Cliffs, 1981.

[52] L. Giraud, S. Gratton, and J. Langou. Convergence in backward error of relaxed GMRES. *SIAM Journal on Scientific Computing*, 29(2):710–728, 2007.

[53] L. Giraud, S. Gratton, X. Pinel, and X. Vasseur. Flexible GMRES with deflated restarting. *SIAM J. Sci. Comput.*, 32(4):1858–1878, 2008.

[54] G. H. Golub and C. F. van Loan. *Matrix Computations.* Johns Hopkins University Press, 3rd edition, 1996.

[55] G. H. Golub and Q. Ye. Inexact preconditioned conjugate gradient method with inner-outer iteration. *SIAM J. Sci. Comput.*, 21:1305–1320, 1999.

[56] S. Goossens and D. Roose. Ritz and harmonic Ritz values and the convergence of FOM and GMRES. *Numerical linear algebra with applications*, 6(4):281–293, 1999.

[57] N. Gould and J. Scott. On approximate-inverse preconditioners. Technical Report RAL-95-026, Rutherford Appleton Laboratory, Chilton, 1995.

[58] S. Gratton. Notes de cours. Unpublished lecture notes, 2008.

[59] A. Greenbaum. *Iterative Methods for Solving Linear Systems*. SIAM, Philadelphia, 1997.

[60] A. Greenbaum, V. Pták, and Z. Strakoš. Any nonincreasing convergence curve is possible for GMRES. *SIAM Journal on Matrix Analysis and Applications*, 17(3):465–469, July 1996.

[61] M. J. Grote and T. Huckle. Parallel preconditioning with sparse approximate inverses. *SIAM Journal on Scientific Computing*, 18(3):838–853, 1997.

[62] G.-D. Gu and H.-B. Wu. A block EN algorithm for nonsymmetric linear systems with multiple righ-hand sides. *Linear Algebra and its Applications*, 299:1–20, 1999.

[63] A. E. Guennouni, K. Jbilou, and H. Sadok. A block version of BICGSTAB for linear systems with multiple right-hand sides. *Electronic Transactions on Numerical Analysis*, 16(1):129–142, 2003.

[64] W. Hackbusch. *Iterative Solution of Large Sparse Linear Systems of Equations*. Springer, 1994.

[65] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, 2nd edition, 2002.

[66] K. Hoffman and R. Kunze. *Linear Algebra*. Prentice-Hall, New Jersey, 2nd edition, 1971.

[67] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1987.

[68] A. S. Householder. *Theory of Matrices in Numerical Analysis*. Dover, 1964.

[69] Intel. www.intel.com. visitado em 30 de março de 2008.

[70] I. Ipsen and C. Meyer. The idea behind Krylov methods. *Amer. Math. Monthly*, 105(10):889–899, 1998.

[71] Z. Jia and G. W. Stewart. An analysis of the Rayleigh-Ritz method for approximating eigenspaces. *Mathematics of Computation*, 70(234):637–647, 2001.

[72] W. Joubert. A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems. *SIAM Journal on Scientific Computing*, 15(2):427–439, 1994.

[73] L. V. Kantorovich and V. I. Krylov. *Approximate methods of higher analysis*. John Wiley & Sons, Inc., 1958.

[74] A. N. Krylov. On the numerical solution of the equation by which in technical questions frequencies of small oscillations of material systems are determined. *News of Academy of Sciences of the USSR*, VII(4):491–539, 1931. (in Russian).

[75] C. Lanczos. *The Variational Principles of Mechanics*. Number 4 in Mathematical Expositions. University of Toronto Press, Toronto, 1948.

[76] S. Lang. *Álgebra Linear*. Editora Ciencia Moderna, 2003.

[77] J. Langou. *Iterative methods for solving linear systems with multiple right hand sides*. Ph.D. dissertation, INSA Toulouse, June 2003. CERFACS-TH/PA/03/24.

[78] T. P. A. Mathew. *Domain Decomposition Methods for the Numerical Solution of Partial Differential Equations*. Springer, 2008.

[79] K. Mer-Nkonga and F. Collino. The fast multipole method applied to a mixed integral system for time-harmonic Maxwell's equations. In B. Michielsen and F. Decavèle, editors, *Proceedings of the European Symposium on Numerical Methods in Electromagnetics*, pages 121–126, Toulouse, France, 2002. ONERA.

[80] G. Meurant. *Computer solution of large linear systems*. North-Holland, 1999.

[81] C. D. Meyer. *Matrix analysis and applied liner algebra*. SIAM, 2000.

[82] R. B. Morgan. Computing interior eigenvalues of large matrices. *Lin. Alg. and Its Applic.*, 154/156:289–309, 1991.

[83] R. B. Morgan. A restarted GMRES method augmented with eigenvectors. *SIAM Journal on Matrix Analysis and Applications*, 16(4):1154–1171, 1995.

[84] R. B. Morgan. Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1112–1135, 2000.

[85] R. B. Morgan. GMRES with deflated restarting. *SIAM Journal on Scientific Computing*, 24(1):20–37, 2002.

[86] N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen. How fast are nonsymmetric matrix iterations? *SIAM Journal on Matrix Analysis and Applications*, 13(3):778–795, 1992.

[87] R. A. Nicolaides. Deflation of conjugate gradients with applications to boundary value problems. *SIAM Journal on Numerical Analysis*, 24(2):355–365, 1987.

[88] Y. Notay. Flexible conjugate gradients. *SIAM Journal on Scientific Computing*, 22(4):1444–1460, 2000.

[89] D. O'Leary. The block conjugate gradient algorithm and related methods. *Linear Algebra and Its Applications*, 29:293–322, 1980.

[90] C. C. Paige, B. N. Parlett, and H. A. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numerical Linear Algebra with Applications*, 2(2):115 – 133, 1995.

[91] C. C. Paige, M. Rozložník, and Z. Strakoš. Modified Gram-Schmidt (MGS), least squares, and backward stability of MGS-GMRES. *SIAM Journal on Matrix Analysis and Applications*, 28(1):264–284, 2006.

[92] M. L. Parks, E. de Sturler, G. Mackey, D. D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems. *SIAM Journal on Scientific Computing*, 28(5):1651–1674, 2006.

[93] B. N. Parlett. *The symmetric eigenvalue problem*. SIAM, Philadelphia, 1998. Corrected reprint of the 1980 original.

[94] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*. Springer-Verlag, 2nd edition, 2007.

[95] M. Robbé and M. Sadkane. Exact and inexact breakdowns in the block GMRES method. *Linear Algebra and its Applications*, 419:265–285, 2006.

[96] M. Rozložník. *Numerical Stability of the GMRES Method*. PhD thesis, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, April 1997.

[97] J. W. Ruge and K. Stüben. *Multigrid Methods*, volume 3 of *Frontiers in Applied Mathematics*, chapter Algebraic multigrid (AMG), pages 73–130. SIAM, Philadelphia, PA, 1987.

[98] Y. Saad. Krylov subspace methods for solving large unsymmetric linear systems. *Math. Comp.*, 37(155):105–126, July 1981.

[99] Y. Saad. A flexible inner-outer preconditioned GMRES algorithm. *SIAM Journal on Scientific Computing*, 14(2):461–469, 1993.

[100] Y. Saad. Analysis of augmented Krylov subspace methods. *SIAM Journal on Matrix Analysis and Applications*, 18(2):435–449, 1997.

[101] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2nd edition, 2003.

[102] Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7(3):856–869, 1986.

[103] Y. Saad and H. A. van der Vorst. Iterative solution of linear systems in the 20th century. *Journal of Computational and Applied Mathematics*, 123(1-2):1–23, 2000.

[104] Y. Saad, M. Yeung, J. Erhel, and F. Guyomarc'h. A deflated version of the conjugate gradient algorithm. *SIAM Journal on Scientific Computing*, 21(5):1909–1926, 2000.

[105] M. Sala, M. A. Heroux, and D. M. Day. Trilinos Tutorial. Technical Report SAND2004-2189, Sandia National Laboratories, 2004.

[106] H. Schwarz. Über eine grenzübergang durch alternirendes verfahren. *Gesammelete Mathematische Abhandlungen, Springer-Verlag*, 2:133–143, 1890. First published in Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, vol.15, pp. 272-286, 1870.

[107] D. S. Scott. The advantages of inverted operators in Rayleigh–Ritz approximations. *SIAM Journal on Scientific and Statistical Computing*, 3(1):68–75, 1982.

[108] J. R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain edition $1\frac{1}{4}$. Technical report, School of Computer Science, Carnegie Mellon University, August 1994.

[109] V. Simoncini. A stabilized QMR version of block BiCG. *SIAM Journal on Matrix Analysis and Applications*, 18(2):419–434, 1997.

[110] V. Simoncini. On the convergence of restarted Krylov subspace methods. *SIAM Journal on Matrix Analysis and Applications*, 22(2):430–452, 2000.

[111] V. Simoncini and D. B. Szyld. Flexible inner-outer Krylov subspace methods. *SIAM Journal on Numerical Analysis*, 40(6):2219–2239, 2002.

[112] V. Simoncini and D. B. Szyld. Theory of inexact Krylov subspace methods and applications to scientific computing. *SIAM Journal on Scientific Computing*, 25(2):454–477, 2003.

[113] V. Simoncini and D. B. Szyld. The effect of non-optimal bases on the convergence of Krylov subspace methods. *Numerische Mathematik*, 100(4):711–733, June 2005.

[114] V. Simoncini and D. B. Szyld. Recent computational developments in Krylov subspace methods for linear systems. *Numerical Linear Algebra with Applications*, 14:1 – 59, 2007.

[115] G. L. G. Sleijpen and J. van den Eshof. On the use of harmonic Ritz pairs in approximating internal eigenpairs. *Linear Algebra and its Applications*, 358(1-3):115–137, January 2003.

[116] G. L. G. Sleijpen and H. A. van der Vorst. A Jacobi–Davidson iteration method for linear eigenvalue problems. *SIAM Review*, 42(2):267–293, 2000. This paper originally appeared in SIAM Journal on Matrix Analysis and Applications, Volume 17, Number 2, 1996, pages 401-425.

[117] B. Smith, P. Bjørstad, and W. Gropp. *Domain Decomposition, Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, New York, 1st edition, 1996.

[118] G. W. Stewart. *Introduction to Matrix Computations*. Academic Press, New York, 1973.

[119] G. W. Stewart. *Matrix Algorithms. Volume I: Basic Algorithms*. SIAM, Philadelphia, PA, USA, 1998.

[120] G. W. Stewart. *Matrix Algorithms. Volume II: Eigensystems*. SIAM, 2001.

[121] G. W. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, 1990.

[122] G. Strang. *Linear Algebra and Its Applications*. Harcourt Brace Jovanovich, Publishers, San Diego, 1988.

[123] D. B. Szyld and J. A. Vogel. FQMR: A flexible quasi-minimal residual method with inexact preconditioning. *SIAM Journal on Scientific Computing*, 23(2):363–380, 2001.

[124] The Mathworks, Inc. *Matlab 7.0*, 2004.

[125] A. Toselli and O. Widlund. *Domain Decomposition methods - Algorithms and Theory*, volume 34 of *Computational Mathematics*. Springer, 2004.

[126] L. N. Trefethen. *Algorithms for Approximation II*, chapter Approximation theory and numerical linear algebra, pages 336–360. Chapman and Hall, London, 1990.

[127] L. N. Trefethen and D. Bau, III. *Numerical linear algebra*. SIAM, 1997.

[128] U. Trottenberg, C. Oosterlee, and A. Schuller. *Multigrid*. Academic Press, 2001.

[129] J. van den Eshof and G. L. G. Sleijpen. Inexact Krylov subspace methods for linear systems. *SIAM Journal on Matrix Analysis and Applications*, 26(1):125–153, 2004.

[130] J. van den Eshof, G. L. G. Sleijpen, and M. B. van Gijzen. Relaxation strategies for nested Krylov methods. *J. Comput. Appl. Math.*, 177:347–365, 2005.

[131] H. van der Vorst. *Iterative Krylov Methods for Large Linear systems*. Cambridge University Press, Cambridge, April 2003.

[132] H. van der Vorst and C. Vuik. GMRESR: a family of nested GMRES methods. *Numerical Linear Algebra with Applications*, 4(1):369–386, 1994.

[133] H. A. van der Vorst and C. Vuik. The superlinear convergence behaviour of GMRES. *J. Comput. Appl. Math.*, 48:327–341, 1993.

[134] C. Vuik. New insights in GMRES-like methods with variable preconditioners. *J. Comput. Appl. Math.*, 61:189–204, 1995.

[135] J. S. Warsa, M. Benzi, T. A. Warein, and J. E. Morel. Preconditioning a mixed continuous finite element method for radiation diffusion. *Numerical Lin. Alg. and Its Applic.*, 11:795–811, 2004.

[136] P. Wesseling. *An Introduction to Multigrid Methods*. Wiley, New York, 1992.

[137] P. Wesseling and C. Oosterlee. Geometric multigrid with applications to computational fluid dynamics. *Journal of Computational and Applied Mathematics*, 128(1-2):311–334, 2001.

[138] J. H. Wilkinson. *Rounding Errors in Algebraic Processes*. Notes on Applied Science No. 32, Her Majesty's Stationery Office, London, 1963. Also published by Prentice-Hall, Englewood Cliffs, NJ, USA. Reprinted by Dover, New York, 1994.

[139] K. Wu and H. Simon. Thick-restart Lanczos method for large symmetric eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications*, 22(2):602–616, 2000.

[140] R. Yu, E. de Sturler, and D. D. Johnson. A block iterative solver for complex non-Hermitian systems applied to large-scale, electronic-structure calculations. 1999.

[141] J.-P. M. Zemke. *Krylov Subspace Methods in Finite Precision: A Unified Approach*. PhD thesis, Technischen Universität Hamburg-Harburg, 2003.

[142] J. Zhang. Sparse approximate inverse and multilevel block ILU preconditioning techniques for general sparse matrices. *Applied Numerical Mathematics*, 35(1):67–86, Sept. 2000.

Índice